

UNIVERSITÉ DE MONTRÉAL

CHERCHEURS UNIVERSITAIRES, FINANCEMENT ET INNOVATION:  
COMPARAISON ENTRE NANOTECHNOLOGIE ET BIOTECHNOLOGIE

RAMINE KANANIAN

DÉPARTEMENT DE MATHÉMATIQUES ET GÉNIE INDUSTRIEL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION  
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES  
(GÉNIE INDUSTRIEL)

DÉCEMBRE 2011

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

CHERCHEURS UNIVERSITAIRES, FINANCEMENT ET INNOVATION:  
COMPARAISON ENTRE NANOTECHNOLOGIE ET BIOTECHNOLOGIE

présenté par : KANANIAN Ramine

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. AGARD Bruno, Doct., président

Mme BEAUDRY Catherine, ing. D.Phil., membre et directrice de recherche

M. LARIVIÈRE Vincent, Ph.D., membre

## REMERCIEMENTS

Je tiens tout d'abord à remercier ma directrice de recherche, Catherine Beaudry, de m'avoir guidé tout au long de ce projet et de m'avoir transmis des connaissances précieuses. Je remercie Vincent Larivière pour l'accès aux données de financement de SIRU et de citations de brevets de la USPTO et pour les informations concernant le degré d'application des journaux. Le logiciel créé par Ahmad Barirani a été d'une grande importance pour l'extraction des données de brevetage de l'USPTO. Je remercie Carl St-Pierre pour son assistance cruciale en ce qui concerne la compilation et l'interprétation des statistiques. Gregory de Crescenzo et Bruno Agard sont aussi à remercier pour leur aide dans la validation des données et la validation de la méthodologie respectivement. Finalement, j'offre ma gratitude à Sedki Allaoui pour son travail impeccable sur la base de données principale et sur le logiciel Pajek et à Benoît Forêt pour l'installation du serveur MySQL.

Je remercie ma directrice pour le soutien financier qu'elle m'a octroyé durant mes études de deuxième cycle. Je suis aussi reconnaissant au Département de Mathématiques et de Génie Industriel et au CIRST pour les bourses de voyage qu'ils m'ont accordées. Leurs contributions m'ont permis d'assister aux conférences de l'ACFAS à l'Université de Sherbrooke et du *Triple Helix* à Stanford University.

## RÉSUMÉ

L'implication des milieux industriels, académiques et gouvernementaux et leur collaboration sont devenues indispensables pour le développement de l'innovation en nanotechnologie et en biotechnologie. Gardant l'importance de ces liens en tête, notre étude évalue la portée du financement public et privé des chercheurs universitaires pour l'innovation en utilisant les flux monétaires provenant du *Système d'information sur la recherche universitaire* (SIRU) et des données de brevets extraites de la *United States Patent and Trademark Office* (USPTO). Nous trouvons que les fonds octroyés, peu importe leur source, n'ont pas une grande incidence sur le nombre d'inventions brevetées. De plus, puisque le modèle d'innovation employé dans la biotechnologie est souvent utilisé comme référence pour innover en nanotechnologie, nous abordons une comparaison statistique de diverses mesures pour vérifier leur similarité. Les éléments compilés sont, entre autres, les intervalles de temps entre le dépôt et l'octroi du brevet, le nombre de chercheurs par brevet, les moyennes d'investissement financier par chercheur, le nombre de revendications, le nombre de citations, les moyennes des degrés d'application, etc. Le tout nous permet de voir que malgré certaines similarités, la nanotechnologie et la biotechnologie n'ont pas autant en commun que ce que l'on pourrait croire. Par ailleurs, nos résultats suggèrent que certains des facteurs définissant la qualité des brevets sont contradictoires. Nous menons une analyse économétrique pour évaluer l'impact du financement sur le brevetage et explorer cette divergence des mesures de qualité de brevets pour les inventeurs universitaires. Le tout se fait en tenant compte de la position du chercheur dans le réseau de collaboration académique, de ses affiliations universitaires et de ses caractéristiques personnelles. Nous trouvons que certains éléments favorisent le brevetage de la part d'universitaires en nanotechnologie et d'autres favorisent le brevetage en biotechnologie. Aussi, la qualité déterminée par le nombre de revendications associées à un brevet est dépendante de différents facteurs de celle associée au nombre de citations.

## ABSTRACT

The involvement of industry, academia and government and their interactions have become of great importance for innovation development in both nanotechnology and biotechnology. Keeping this in mind, using funding data provided by the *Système d'information sur la recherche universitaire* (SIRU) and patenting data extracted from the *United States Patent and Trademark Office* (USPTO), our study evaluates the importance of private and public funding of university researchers for innovation. We find that awarded funds, whether they are of a public or private origin, do not have much of an effect on subsequent patenting. Furthermore, since the innovation model used in biotechnology is often taken as a reference for efficient innovation in nanotechnology, we compile comparison statistics of various metrics to verify their similarity. Some of the compiled measures include: time intervals between patent application and granting, the number of researchers per patent, the average financial investment per researcher, the number of claims associated to each patent, the number of forward citations, the degree of application, etc. This enables us to see that despite their similarities, nanotechnology and biotechnology do not have as much in common as one might think. Moreover our results suggest contradictory findings as far as quality indicators are concerned, underlining a potential unreliability of one or the other. We conduct an econometric analysis to evaluate the impact of funding on university researcher patenting and to obtain better insight into this divergence of patent quality indicators. In doing this, we also take into account the researcher's position in the academic network, university affiliations and other personal characteristics. We find that certain factors increase innovation output of university researchers in nanotechnology, while others are more useful for innovation in biotechnology. Also the quality of a patent as defined by its claims depends on certain key points that differ from the quality associated with its citation rate.

## TABLE DES MATIÈRES

REMERCIEMENTS .....	iii
RÉSUMÉ.....	iv
ABSTRACT .....	v
TABLE DES MATIÈRES .....	vi
LISTE DES TABLEAUX.....	ix
LISTE DES FIGURES.....	xv
LISTE DES SIGLES ET ABRÉVIATIONS .....	xviii
LISTE DES ANNEXES.....	xx
INTRODUCTION.....	1
1.1    Contexte .....	1
1.2    Problématique.....	2
1.3    But de l'étude et objectifs.....	2
1.4    Organisation de la recherche .....	3
CHAPITRE 2    RECENSION DE LA LITTÉRATURE.....	4
2.1    Domaines à l'étude.....	4
2.1.1    Nanotechnologie.....	4
2.1.2    Biotechnologie .....	8
2.1.3    Note sur la nanobiotechnologie .....	10
2.2    Innovation.....	10
2.2.1    Modèle d'innovation .....	11
2.2.2    Financement de chercheurs universitaires et innovation.....	15
2.2.3    Réseaux .....	17
2.2.4    Fondement de la comparaison de la nanotechnologie et de la biotechnologie .....	18

2.3	Synthèse .....	19
CHAPITRE 3 DONNÉES, HYPOTHÈSES ET MÉTHODOLOGIE.....		20
3.1	Hypothèses .....	20
3.2	Objectifs .....	23
3.3	Données .....	24
3.3.1	Financement .....	24
3.3.2	Brevetage.....	25
3.3.3	Publication.....	26
3.4	Méthodologie .....	27
3.4.1	Indices scientométriques et mesures .....	27
3.4.2	Identification et assignation des noms de chercheurs .....	31
3.4.3	Identification et assignation des cessionnaires.....	32
3.4.4	Bouclage.....	33
3.4.5	Degré d'application .....	34
3.4.6	Classification des brevets en sous-catégories.....	35
3.4.7	Méthodologie Statistique Comparative .....	44
3.4.8	Méthodologie Économétrique .....	49
CHAPITRE 4 RÉSULTATS ET ANALYSE.....		72
4.1	Similarité en nanotechnologie et biotechnologie .....	72
4.2	Processus de bouclage.....	74
4.3	Qualité de brevet .....	84
4.4	Degré d'application .....	85
4.5	Régressions.....	86
4.5.1	Analyse du bloc A - breveter et nombre de brevets .....	87

4.5.2	Comparaison entre les blocs B et C - breveter avec revendications et citations .....	89
4.5.3	Analyse du bloc D – nombre de brevets .....	91
4.5.4	Comparaison entre les blocs E et F – nombre de revendications et de citations.....	92
CHAPITRE 5 DISCUSSION GÉNÉRALE .....		95
CONCLUSION ET RECOMMANDATIONS .....		98
5.1	Synthèse .....	98
5.2	Limitations .....	98
5.3	Implications sur la politique .....	100
5.4	Directions de recherche .....	100
BIBLIOGRAPHIE .....		102
RÉFÉRENCES .....		102
ANNEXES .....		112



## LISTE DES TABLEAUX

Tableau 3-1: Précision de la classification selon la méthode d'apprentissage .....	40
Tableau 3-2 : matrice de corrélation entre la variable endogène et les variables instrumentales de [L.itotsubvmoyan3] .....	61
Tableau 3-3 : matrice de corrélation entre la variable dépendante et les variables instrumentales	62
Tableau 3-4: Liste des régressions finales effectuées et des procédures de Stata utilisées pour chacun des modèles .....	65
Tableau 3-5 : Régressions du bloc A sur le fait de breveter combiné au nombre de brevets .....	66
Tableau 3-6: Régressions du bloc B sur le nombre de revendications combinant le fait de breveter au nombre comme tel .....	67
Tableau 3-7 : Régressions du bloc C sur le nombre de citations combinant le fait de breveter au nombre comme tel .....	68
Tableau 3-8 : Régressions du bloc D sur le nombre de brevets si et seulement si le chercheur possède au moins un brevet.....	69
Tableau 3-9: Régressions du bloc E sur le nombre de revendications si et seulement si le chercheur possède au moins un brevet.....	70
Tableau 3-10: Régressions du bloc F sur le nombre de citations si et seulement si le chercheur possède au moins un brevet.....	71
Tableau 4-1 : Qualité des brevets et degré d'application .....	73
Tableau 4-2 : Degré de signification des comparaisons de qualité et du degré d'application pour $H_0$ : égalité des moyennes.....	73
Tableau 4-3 : Statistiques sommaires des boucle d'innovation .....	75
Tableau 4-4 : Statistiques descriptives de boucles d'innovation en nanotechnologie selon les sous-domaines .....	76
Tableau 4-5: Statistiques descriptives de boucles d'innovation en biotechnologie selon les sous-domaines.....	77

Tableau 4-6 : Indicateurs de qualité des brevets appartenant à des firmes .....	84
Tableau 4-7: Degré d'application de connaissances associées aux brevets en boucles selon l'intervalle de temps entre le financement et le dépôt de brevet .....	85
Tableau A-1 : Table SIRU .....	112
Tableau A-2 : Données ajoutées à la table SIRU .....	113
Tableau A-3 : Information générale sur les brevets .....	113
Tableau A-4 : Information sur le type de brevet .....	114
Tableau A-5 : Information sur les cessionnaires de brevets.....	114
Tableau A-6 : Information sur les inventeurs de brevets .....	114
Tableau D-1: Inconvénients d'utiliser le forage de données dans l'évaluation des boucles d'innovation .....	124
Tableau D-2: Désavantages d'effectuer une révision individuelle des boucles d'innovation .....	125
Tableau I-1: Précision de la classification k-nn des nanoprocessus et nanométrie en nanotechnologie .....	191
Tableau I-2: Performance par classification k-nn des nanoprocessus et nanométrie en nanotechnologie .....	192
Tableau I-3: Précision de la classification Naive Bayes des nanoprocessus et nanométrie en nanotechnologie .....	193
Tableau I-4: Performance par classification Naive Bayes des nanoprocessus et nanométrie en nanotechnologie .....	193
Tableau I-5: Précision de la classification SVM des nanoprocessus et nanométrie en nanotechnologie .....	194
Tableau I-6: Performance par classification SVM des nanoprocessus et nanométrie en nanotechnologie .....	194
Tableau I-7: Précision de la classification k-nn de la nanoélectronique et des nanodispositifs en nanotechnologie .....	195

Tableau I-8: Performance par classification k-nn de la nanoélectronique et des nanodispositifs en nanotechnologie .....	196
Tableau I-9: Précision de la classification Naive Bayes de la nanoélectronique et des nanodispositifs en nanotechnologie .....	197
Tableau I-10: Performance par classification Naive Bayes de la nanoélectronique et des nanodispositifs en nanotechnologie .....	197
Tableau I-11: Précision de la classification SVM de la nanoélectronique et des nanodispositifs en nanotechnologie .....	198
Tableau I-12: performance par classification SVM de la nanoélectronique et des nanodispositifs en nanotechnologie.....	198
Tableau I-13: Précision de la classification k-nn des nanomatériaux et des nanostructures en nanotechnologie .....	199
Tableau I-14: Performance par classification k-nn des nanomatériaux et des nanostructures en nanotechnologie .....	200
Tableau I-15: Précision de la classification Naive Bayes des nanomatériaux et des nanostructures en nanotechnologie.....	201
Tableau I-16: Performance par classification Naive Bayes des nanomatériaux et des nanostructures en nanotechnologie .....	201
Tableau I-17: Précision de la classification SVM des nanomatériaux et des nanostructures en nanotechnologie .....	202
Tableau I-18: Performance par classification SVM des nanomatériaux et des nanostructures en nanotechnologie .....	202
Tableau J-1: Précision de la classification k-nn en biotechnologie .....	204
Tableau J-2: Performance par classification k-nn par valeur de k .....	204
Tableau J-3: Précision de la classification Naïve Bayes en biotechnologie .....	205
Tableau J-4: Performance par classification Naive Bayes par correction de Laplace en biotechnologie .....	205

Tableau J-5: Précision de la classification SVM en biotechnologie .....	206
Tableau J-6: Performance par classification SVM par type de noyau en biotechnologie.....	206
Tableau L-1: Définition de la biotechnologie selon les techniques employées. Source: OECD(2005). A Framework for Biotechnology Statistics. OECD, Paris. ....	208
Tableau L-2: Classification de la biotechnologie selon les applications. Source: OECD(2005). A Framework for Biotechnology Statistics. OECD, Paris. ....	209
Tableau L-3: Catégorisation de brevets selon le code de classification de contenu de la base de données Derwent. Source: Dalpé (2002). Bibliometric Analysis of Biotechnology. ....	210
Tableau L-4: Catégorisation de brevets selon le code de classification de contenu de la base de données Derwent (Abrégé). Source : Ganerjee, Gupta & Garg (2000). Patent Statistics as indicators of competition, an analysis of patenting in biotechnology. NISTADS, New Delhi. ....	211
Tableau L-5: Secteurs industriels de la biotechnologie. Source : Traoré (2004). Biotechnology Use and Development Survey : methodology, issues and responses. Statistics Canada. Ottawa. ....	212
Tableau L-6 : Secteurs d'activités de la Biotechnologie et agrégation. Source : Arundel (1999). Diffusion of biotechnologies in Canada: Results from the survey of biotechnology use in Canadian Industries. Statistics Canada.....	213
Tableau L-7: Pourcentage d'utilisation des biotechnologies selon catégorisation en secteurs et en applications.....	213
Tableau M-1: table de corrélation entre des variables pour les blocs A, B et C .....	214
Tableau M-2: table de corrélation entre des variables pour les blocs D, E et F.....	215
Tableau N-1: Régressions à binomiale négative (deuxième étape) du bloc A – fait de breveter et nombre de brevets .....	216
Tableau N-2: Régressions à binomiale négative en panel (deuxième étape) du bloc A - fait de breveter et nombre de brevets .....	219

Tableau N-3: Régressions à binomiale négative à zéro-augmenté (deuxième étape) du bloc A – fait de breveter et nombre de brevets .....	222
Tableau N-4: Régressions tobit (deuxième étape) du bloc A – fait de breveter et nombre de brevets .....	224
Tableau O-1: Régressions à binomiale négative (deuxième étape) du bloc B – breveter avec revendications.....	226
Tableau O-2: Régressions à binomiale négative en panel (deuxième étape) du bloc B – breveter avec revendications .....	229
Tableau O-3: Régressions à binomiale négative à zéros-augmenté (deuxième étape) du bloc B – breveter avec revendications .....	232
Tableau O-4: Régressions tobit (deuxième étape) du bloc B – breveter avec revendications .....	234
Tableau P-1: Régressions à binomiale négative (deuxième étape) du bloc C – breveter avec citations .....	236
Tableau P-2: Régressions à binomiale négative en panel (deuxième étape) du bloc C – breveter avec citations .....	238
Tableau P-3: Régressions à binomiale négative à zéro-augmenté (deuxième étape) du bloc C – breveter avec citations .....	241
Tableau P-4: Régressions tobit (deuxième étape) du bloc C – breveter avec citations .....	243
Tableau Q-1: Régressions à binomiale négative (deuxième étape) du bloc D – nombre de brevets .....	245
Tableau Q-2: Régressions tobit (deuxième étape) du bloc D – nombre de brevets .....	247
Tableau R-1: Régressions à binomiale négative (deuxième étape) du bloc E – nombre de revendications.....	249
Tableau R-2: Régressions à Moindres Carrés Ordinaires (deuxième étape) du bloc E – nombre de revendications.....	252
Tableau R-3: Régressions à binomiale négative (deuxième étape) du bloc E – nombre de revendications avec sous-catégories de brevets .....	254

Tableau R-4: Régressions à Moindres Carrés Ordinaires (deuxième étape) du bloc E – nombre de revendications avec sous-catégories de brevets .....	257
Tableau S-1: Régressions à binomiale négative (deuxième étape) du bloc F – nombre de citations .....	260
Tableau S-2: Régressions tobit sans endogénéité et de deuxième étape du bloc F – nombre de citations .....	263
Tableau S-3: Régressions à binomiale négative (deuxième étape) du bloc F – nombre de citations avec sous-catégories de brevets.....	265
Tableau S-4: Régressions Tobit (deuxième étape) du bloc F - nombre de citations avec sous-catégories de brevets .....	268
Tableau T-1: Régressions de première étape en Moindres Carrés Ordinaires pour le bloc A – breveter et nombre de brevets .....	270
Tableau T-2: Régressions de première étape en Moindre Carrés Ordinaires en panel pour le bloc A – breveter et nombre de brevets .....	272
Tableau T-3: Régressions de première étape en Moindre Carrés Ordinaires pour les blocs B et C – breveter avec revendications et citations.....	274
Tableau T-4: Régressions de première étape en Moindres Carrés Ordinaires en panel pour les blocs B et C – breveter avec revendications et citations .....	276
Tableau U-1: Régressions de première étape en Moindres Carrés Ordinaires pour le bloc D – nombre de brevets .....	278
Tableau U-2: Régressions de première étape en Moindres Carrés Ordinaires pour les blocs E et F – nombre de revendications et citations .....	280
Tableau W-1: Liste des variables employées .....	285
Tableau W-2: Liste des variables employées (suite).....	287

## LISTE DES FIGURES

Figure 2-1: Deux types de base de transfert de connaissances et de technologie à la commercialisation (Butler, 2011).....	11
Figure 2-2: Modèle Triple (Quadruple) Hélice .....	13
Figure 3-1 : Boucle d'innovation .....	21
Figure 3-2: Méthode de catégorisation des brevets dans Rapidminer .....	41
Figure 3-3: Méthode employée dans l'optimisation des paramètres et la combinaison des classifications dans Rapidminer .....	42
Figure 3-4: Classification de brevets nanotechnologiques.....	43
Figure 3-5: Classification de brevets biotechnologiques .....	43
Figure 3-6: Arbre décisionnel pour tester les différences de moyennes et variances .....	48
Figure 3-7: Arbre décisionnel des régressions pour les données de comptage. ....	55
Figure 4-1: Proportion de brevets qui ont au moins un chercheur contenu dans une boucle.....	79
Figure 4-2: Charte de collaboration organisation-scientifique en biotechnologie (nombre de boucles d'innovation par partenariat).....	81
Figure 4-3: Charte de collaboration organisation-scientifique en nanotechnologie (nombre de boucles d'innovation par partenariat).....	82
Figure 4-4: Proportion de brevets ayant des chercheurs universitaires comme inventeurs .....	83
Figure B-1: Financement de chercheurs québécois en nanotechnologie .....	115
Figure B-2 : Financement de chercheurs québécois en biotechnologie .....	116
Figure B-3 : Top des détenteurs de brevets en biotechnologie .....	116
Figure B-4: Top des détenteurs de brevets en nanotechnologie.....	116
Figure B-5: Meilleurs inventeurs en biotechnologie.....	117
Figure B-6: Meilleurs inventeurs en nanotechnologie .....	117

Figure B-7: Évolution du brevetage en nanotechnologie issu d'inventions de chercheurs québécois.....	118
Figure B-8: Évolution du brevetage en biotechnologie issu d'inventions de chercheurs québécois .....	118
Figure B-9: Évolution du brevetage en nanobiotechnologie issu d'inventions de chercheurs québécois .....	119
Figure E-1: Average degree of application of knowledge associated to patented inventions.....	126
Figure E-2: Number of patents derived from loops .....	126
Figure I-1: Optimisation du paramètre k de la classification k-nn des nanoprocessus et nanométrie en nanotechnologie.....	191
Figure I-2 : Optimisation du paramètre de la correction de Laplace de la classification Naive Bayes des nanoprocessus et nanométrie en nanotechnologie .....	193
Figure I-3: Optimisation du paramètre de type de noyau de la classification SVM des nanoprocessus et nanométrie en nanotechnologie .....	194
Figure I-4: Optimisation du paramètre k de la classification k-nn de la nanoélectronique et des nanodispositifs en nanotechnologie .....	195
Figure I-5: Optimisation du paramètre de la correction de Laplace de la classification Naive Bayes de la nanoélectronique et des nanodispositifs en nanotechnologie .....	197
Figure I-6: Optimisation du paramètre de type de noyau de la classification SVM de la nanoélectronique et des nanodispositifs en nanotechnologie.....	198
Figure I-7: Optimisation du paramètre k de la classification k-nn des nanomatériaux et des nanostructures en nanotechnologie .....	199
Figure I-8: Optimisation du paramètre de la correction de Laplace de la classification Naive Bayes des nanomatériaux et des nanostructures en nanotechnologie .....	201
Figure I-9: Optimisation du paramètre k de la classification k-nn des nanomatériaux et des nanostructures en nanotechnologie .....	202
Figure J-1: Optimisation du paramètre k de la classification k-nn de biotechnologie .....	203



Figure J-2: Optimisation du paramètre de la correction de Laplace(vrai ou faux) de la classification Naive Bayes en biotechnologie.....	205
Figure J-3: Optimisation du paramètre de type de noyau de la classification SVM de biotechnologie .....	206
Figure K-1 : Diagramme des sous-sections de la nanotechnologie. Source : Porter et al. (2007). Refining search terms for nanotechnology.....	207
Figure V-1: Extrait des résultats de la classification dans la catégorie de nanostructures et nanomatériaux .....	282
Figure V-2: Tests d'Hausman et de Breusch-Pagan dans Stata.....	283
Figure V-3: Test de Durbin-Wu-Hausman dans Stata .....	284

## LISTE DES SIGLES ET ABRÉVIATIONS

La liste des sigles et abréviations présente, dans l'ordre alphabétique, les sigles et abréviations utilisés dans le mémoire ainsi que leur signification.

AIC	Akaike Information Criterion
AFM	Atomic Force Microscope
BIC	Bayesian Information Criterion
CAD	Dollar Canadien
CIHR	Canadian Institutes of Health Research
CIRST	Center for Interuniversity Research on Science and Technology
CNRC	Conseil National de Recherche Canada
DSTI	Directorate for Science Technology and Industry
EPO	European Patent Office
FSU	Former Soviet Union
GT CNS-ASU	Georgia Tech Centre for Nanotechnology in Society at Arizona State University
IC	Intercooled (i.e. standard)
IPC	Indice des Prix à la Consommation
IPP	Incidental Parameter Problem
JPO	Japan Patent Office
K-NN	K-Nearest Neighbors (à ne pas confondre avec K-means)
MELS	Ministère de l'Éducation, du Loisir et du Sport
MBE	Molecular Beam Epitaxy
MCO	Moindres Carrés Ordinaires
MOCVD	Metal-Organic Chemical Vapour Deposition
NBREG	Negative Binomial Regression

NEMS	Nanoelectromechanical Systems
NINT	National Institute for Nanotechnology
NM	Nanomètre
NNI	National Nanotechnology Institute
NSERC	National Sciences and Engineering Research Council
OECD	Organization for Economic Co-operation and Development
CIPO	Office de la Propriété Intellectuelle du Canada
R&D	Recherche et Développement
SCOPUS	Base de données de résumés et de citations d'articles scientifiques
SIRU	Système d'information sur la recherche universitaire
SME	Small and Medium Enterprises
SSHRC	Social Sciences and Humanities Research Council
STM	Scanning Force Microscope
SVM	Support Vector Machine
TERMIUM	Banque de données terminologiques et linguistiques du gouvernement du Canada
TF-IDF	Term Frequency-Inverse Document Frequency
USD	Dollar Américain
USPTO	United States Patent and Trademark Office
XML	Extensible Markup Language
ZINB	Zero-Inflated Negative Binomial Regression

## LISTE DES ANNEXES

- ANNEXE A Tableaux utilisés dans l'analyse des données
- ANNEXE B Tableaux de statistiques descriptives sur le financement et le brevetage
- ANNEXE C Extraits de code SQL
- ANNEXE D Inconvénients de l'utilisation de forage de données et de l'inspection manuelle pour la validation des boucles
- ANNEXE E Graphiques supplémentaires
- ANNEXE F Procédure pour tester différences de moyennes et variances
- ANNEXE G Code XML pour classification de brevets nanotechnologiques en sous-catégories
- ANNEXE H Code XML pour évaluer la performance des systèmes d'apprentissage; cas du Support Vector Machine pour la biotechnologie
- ANNEXE I Performance de la classification de brevets en nanotechnologie selon la méthode (K-NN, Naive Bayes, Support Machine Vector)
- ANNEXE J Performance de la classification de brevets en biotechnologie selon la méthode (K-NN, Naive Bayes, Support Machine Vector)
- ANNEXE K Options de catégorisation de la nanotechnologie
- ANNEXE L Options de catégorisation de la biotechnologie et applications
- ANNEXE M Tables de corrélation
- ANNEXE N Résultats finaux des régressions du bloc A
- ANNEXE O Résultats finaux des régressions du bloc B
- ANNEXE P Résultats finaux des régressions du bloc C
- ANNEXE Q Résultats finaux des régressions du bloc D
- ANNEXE R Résultats finaux des régressions du bloc E
- ANNEXE S Résultats finaux des régressions du bloc F
- ANNEXE T Résultats finaux des régressions de première étape pour les grands échantillons

ANNEXE U Résultats finaux des régressions de première étape pour les petits échantillons

ANNEXE V Extraits procéduraux de forage de texte et de méthodes économétriques

ANNEXE W Variables employées dans les régressions

## INTRODUCTION

### 1.1 Contexte

Au Québec, comme en Amérique du nord, l'industrie de la fabrication en grandes séries est, comme nous le savons tous, une espèce en voie de disparition. La mondialisation durant les trois dernières décennies a non seulement permis, mais aussi facilité, la délocalisation de nombreuses industries telles que le textile, le meuble et l'électronique vers l'Orient et l'Amérique du Sud (Bidanda et al., 2006; Bardan & Kroll, 2003). Il est donc normal que les chaînes de production à *la Ford*, souvent plus onéreuses à garder ouvertes ici qu'à l'étranger, soient de plus en plus rares dans des pays développés, (Tchana, 2009). Ceci est vrai tant au plan provincial que fédéral. Mais comment une société, qui a une production du secteur primaire limitée par son territoire et un secteur secondaire qui laisse à désirer, parvient-elle à maintenir une économie saine et viable? Elle pourrait, bien entendu, baisser les bras et accepter qu'il y ait des pays qui sont mieux munis qu'elle en termes de capacité de production et de main d'œuvre quasi-gratuite, mais ceci ne semble de toute évidence pas être une option saine et viable. La voie qui semble être la plus soutenue aujourd'hui, et depuis quelque temps maintenant dans certaines industries, est celle d'investir en innovation (Godin et Trépanier, 1995).

L'innovation, à travers la recherche et développement, permet ainsi à celui qui l'entreprend de rester à l'avance de ses concurrents, ce qui dans plusieurs cas signifie garantir sa pérennité (Rao et al., 2001). Une industrie bourgeonnante, qui croît au Canada grâce à la R&D qui y est consacrée, est la nanotechnologie (Lafontaine, 2006; Marinova et McAleer, 2002), c'est-à-dire « la technologie où les dimensions ou tolérances dans l'ordre de 0.1 à 100nm (de la taille d'un atome à la longueur d'onde de la lumière) jouent un rôle critique » (Franks, 1987). C'est un domaine de haute importance pour l'économie nationale, puisque les avancements en nanotechnologie sont applicables dans presque tous les domaines, e.g. vêtement, construction, santé et électronique pour n'en nommer que quelques-uns, et révolutionnera les technologies de l'information, les matériaux et la médecine (Meyer & Persson, 1998; Budworth, 1996; Bachmann, 1995; Nelson & Shipaugh, 1995) mais aussi parce que le marché mondial pour les marchandises dans lesquelles elle est utilisée sera de \$ 1,5 trillion USD d'ici 2015 (CNRC, 2010).

## 1.2 Problématique

L'importance de la nanotechnologie réside principalement dans le fait qu'elle est une percée *grilichésienne*, génératrice de progrès métamorphique; en d'autres mots, parce qu'elle est une invention de méthodes d'inventions, elle engendre une croissance de productivité plus rapidement que la moyenne (Darby & Zucker, 2005). Cette fonction de catalyseur économique est aussi mise à l'avant dans un autre domaine un peu moins jeune. Il s'agit de la biotechnologie.

Le transfert de la plateforme d'innovation de la biotechnologie dans le domaine de la nanotechnologie est rendu pratique commune. En effet, plusieurs analyses révélant les similarités entre la nanotechnologie et la biotechnologie (Darby & Zucker, 2003; Zucker *et al.* 1998; Rothaermel & Thursby, 2007; Thompson, 2008) sont utilisées comme outils pour la justifier. De plus, plusieurs analyses comparatives sont menées pour extraire des apprentissages en biotechnologie et les appliquer à la nanotechnologie (MacNaghten, 2008; Sparling, 2008). Cependant, au meilleur de nos connaissances, aucune analyse quantitative n'a été menée pour vérifier cette similarité au Québec. Ceci est d'une importance capitale car, il se pourrait que l'innovation en nanotechnologie ne soit pas aussi similaire que l'on pense à celle en biotechnologie et la transposition de méthode d'innovation d'un domaine dans l'autre serait conséquemment inappropriée.

## 1.3 But de l'étude et objectifs

Le but initial de cette étude est de déceler les similarités et les différences entre le processus d'innovation en nanotechnologie et celui en biotechnologie et pour fins de concision, nous limitons notre analyse à une partie spécifique du processus d'innovation : le financement de chercheurs universitaires et le brevetage qui en découle. Cependant, en cours de recherche nous découvrons qu'il y a une divergence associée au concept de qualité de brevet que nous décidons d'explorer davantage. Ainsi un autre objectif s'ajoute à notre liste, celui d'expliquer l'écart des concepts de qualité que sont le nombre de revendications associées à un brevet et le nombre de citations qu'il reçoit.

## **1.4 Organisation de la recherche**

Le mémoire est organisé comme suit. Le chapitre 2 contient une revue critique de la littérature abordant les sujets de l'innovation dans les domaines de la nanotechnologie et de la biotechnologie, la Triple Hélice et ses composantes, le financement et les extrants de subventions. Le chapitre 3 contient la démarche de l'ensemble du projet de recherche et l'organisation générale décrivant la cohérence des sections par rapport aux objectifs de recherche, les données, les hypothèses et les aspects méthodologiques. Le chapitre 4 contient les résultats et leur analyse. Le chapitre 5 contient la discussion générale qui revient sur les résultats et la validation des hypothèses. Finalement, nous enchaînons avec une conclusion qui souligne les limites de cette recherche, qui met en valeur les implications soulevées sur la politique économique et les directions de recherche potentielles.



## CHAPITRE 2 RECENSION DE LA LITTÉRATURE

Notre objectif dans ce chapitre est d'établir une compréhension des mécanismes existants supportant et facilitant l'innovation dans les industries de haute technologie que sont la nanotechnologie et la biotechnologie. Afin de bien saisir l'étendue de ce qui est à l'étude, nous commençons par la mise au point des définitions et de la portée réelle de chaque champ. Nous abordons ensuite les composantes du modèle d'innovation, l'importance de la collaboration du point de vue d'un pourvoyeur de fonds et du chercheur financé. Finalement, l'évaluation des similarités et différences entre la nanotechnologie et la biotechnologie est ensuite effectuée, le tout dans l'objectif de supporter la formulation des hypothèses de recherche.

### 2.1 Domaines à l'étude

#### 2.1.1 Nanotechnologie

##### 2.1.1.1 Définitions

Qu'est-ce que la nanotechnologie? Une grande gamme de définitions existe (OCDE, 2009) et certaines sont plus inclusives que d'autres. Selon la *National Nanotechnology Institute*

“Nanotechnology is the understanding and control of matter at dimensions between approximately 1 and 100 nanometers, where unique phenomena enable novel applications. Encompassing nanoscale science, engineering, and technology, nanotechnology involves imaging, measuring, modeling, and manipulating matter at this length scale. A nanometer is one-billionth of a meter. A sheet of paper is about 100,000 nanometers thick; a single gold atom is about a third of a nanometer in diameter. Dimensions between approximately 1 and 100 nanometers are known as the nanoscale. Unusual physical, chemical, and biological properties can emerge in materials at the nanoscale. These properties may differ in important ways from the properties of bulk materials and single atoms or molecules”.

Le septième programme cadre de l'Union Européenne décrit la nanotechnologie comme “Generating new knowledge on interface and size-dependent phenomena; nano-scale control of material properties for new applications; integration of technologies at the nano-scale; self-assembling properties; nano-motors; machines and systems; methods and tools for characterization and manipulation at nano-dimensions; nano-precision technologies in chemistry for the manufacture of basic materials and components; impact on human safety, health and the environment; metrology, monitoring and sensing, nomenclature and standards; exploration of

new concepts and approaches for sectorial applications, including the integration and convergence of emerging technologies.” *ISO TC229* définit la nanotechnologie comme étant “Understanding and control of matter and processes at the nano-scale, typically, but not exclusively, below 100 nanometers in one or more dimensions where the onset of size-dependent phenomena usually enables novel applications - utilizing the properties of nanoscale materials that differ from the properties of individual atoms, molecules, and bulk matter, to create improved materials, devices, and systems that exploit these new properties.” Au Japon, La *Second Science and Technology Basic Plan* offre la définition suivante: “Nanotechnology is an interdisciplinary S&T that encompasses IT technology, the environmental sciences, life sciences, materials science, etc. It is for controlling and handling atoms and molecules in the order of nano (1/1 000 000 000) meter enabling discovery of new functions by taking advantage of its material characteristics unique to nano-size, so that it can bring technological innovation in various fields”. Le Royaume-Uni, à travers la *New Dimensions for Manufacturing: A UK Strategy for Nanotechnology*, propose

“Nanotechnology and nanoscience are concerned with materials science and its application at, or around, the nanometer scale (1 billionth of a meter). Manufacturing can reach the nano-scale either from the top down, by ‘machining’ to ever smaller dimensions, or from the bottom up, by exploiting the ability of molecules and biological systems to ‘self-assemble’ tiny structures. It is in the conjunction of these two approaches, in the meeting of physical and chemical/biological manufacturing, that the potential for revolution lies. From the top down perspective, it interfaces with the larger-scale, more mature ‘microsystems technology’ being pursued very actively in the UK and around the world on a more immediate timescale.”

Ainsi, tout dépendant de la définition adoptée, les inférences statistiques peuvent grandement varier. Il est cependant important de mentionner que toutes ces définitions se chevauchent globalement et que de manière générale, la définition donnée par Franks (1987) est relativement bien acceptée par la communauté scientifique (Meyer & Persson, 1998). Il décrit la nanotechnologie comme la technologie où les dimensions et tolérances de 0,1 à 100 nm jouent un rôle critique.

#### **2.1.1.2 État de l’art de la nanotechnologie**

Après avoir recueilli de l’information de diverses sources, Hullman (2006) propose qu’il y aura une croissance significative du marché des produits nanotechnologiques après 2010 et qu’il

englobera près de 15% de la production manufacturière mondiale. Roco (2002) a d'ailleurs estimé que cette production mondiale dans le domaine de la nanotechnologie dépassera 1 trillion USD dans un horizon de 10 à 15 ans. Selon Hullman (2006), tout dépendant de la définition de la nanotechnologie et de sa contribution à la valeur ajoutée de produits finaux, les prévisions du marché global varient entre 150 milliards USD en 2010 (Mistubishi Institute, 2002) et 2.6 trillion USD en 2014 (Lux Research, 2004). Par ailleurs, la NSF estime que 2 millions de travailleurs seront nécessaires pour combler la demande de production d'ici 2015 (Roco, 2001). Une des raisons pour lesquelles les prévisions des parts de marché de la nanotechnologie sont si grandes est le fait qu'elle est reconnue comme technologie clef du XXI siècle et qu'elle révolutionnera plusieurs domaines, dont les technologies de l'information, la médecine, etc. (Meyer & Persson, 1998). Par exemple, la nanoélectronique contribue à la loi de Moore sur le doublage de la mémoire et du traitement (*processing*) à tous les 18 mois (Hullman, 2006) et certains nanomatériaux sont utilisés dans la livraison des médicaments.

Sur le plan global, les grands investisseurs en recherche et développement (R&D) nanotechnologique sont l'Europe de l'Ouest, le Japon, les États-Unis et certains autres (Australia, Canada, China, Europe de l'Est, FSU, Corée, Singapore, Taiwan, etc.) (Roco, 2002). Même si plusieurs pays ont injecté de grandes sommes dans le développement de la nanotechnologie dans les années '80 et '90, Roco (2001) suggère que la nanotechnologie n'a évolué que de la science-fiction à une entité concrète suite à la création du premier programme cohérent de nanotechnologie en janvier 2000 avec la participation de plusieurs agences fédérales américaines clefs et du secteur privé. Ceci a été un événement catalyseur, engendrant la création de plusieurs programmes nanotechnologiques dans pratiquement tous les autres pays développés en 2001, dont le Canada avec le National Institute of Nanotechnology (NINT). Hullman (2007) souligne que selon la Commission Européenne, 37,9 millions € des 3,85 milliards € investis globalement en nanotechnologie en tant que financement public aurait été dépensé en R&D nanotechnologique au Canada en 2005, ce qui le met en 16<sup>ième</sup> place. Ainsi, près de 1% du financement public de la R&D nanotechnologique a été dépensé au Canada à l'année en question. De plus, en terme de production scientifique le Canada est 13<sup>ième</sup> pour le nombre d'articles publiés, avec 1579 articles en 2005 (Kostoff *et al.*, 2007). Il est aussi important de mentionner que mis à part le NINT, le Canada a plusieurs cadres provinciaux pour supporter la R&D en

nanotechnologie : British Columbia Nanotechnology Alliance (Nanotech BC), NanoAlberta, NanoQuébec et la Nanotechnology Network of Ontario (OCDE, 2009).

La nanotechnologie est donc importante tant sur la scène nationale qu'internationale, pour les apports technologiques et économiques qu'elle procure mais aussi pour les ressources qui sont consacrées à son développement.

### 2.1.1.3 Sous-domaines

La nanotechnologie est un domaine d'innovation très vaste qui touche à plusieurs secteurs<sup>1</sup>. Un nombre limité d'études ont été effectuées avec des sous-catégorisations<sup>2</sup> de la nanotechnologie. Pour les fins de notre analyse, nous optons pour l'utilisation de celle de la *Georgia Tech Technology and Assessment Center* (GT CNS-ASU Group). Elle est divisée en 4 catégories définies comme suit :

- (1) Métrologie & nano-processus: Microscopie, i.e. microscopie à sonde locale & microscopie électronique, auto-assemblage, assemblage dirigé, nano-mécanique, simulation moléculaire, fabrication à sonde locale, processus descendant comme la nano-lithographie et la nano-usinage par laser;
- (2) Nano-dispositifs & nanoélectronique : dispositifs de nano-informatique, nano-transistors, nano-systèmes électromécaniques & sondes encapsulées par scellage biologiquement localisés, électronique moléculaire, magnétisme nanométrique;
- (3) Nano-médecine & nanobiotechnologie : dispositifs biomoléculaires, bio-senseurs, moteurs moléculaires, tissus biomoléculaires, enzymes et protéines génétiquement modifiées, libération et découverte de médicaments;
- (4) Nanostructure chimique & nanomatériaux : structures chimiques nanométriques, nano-composites, sol-gel, quasi-cristaux, méthodes de croissance, épitaxie (MBE, CBE,

---

<sup>1</sup> En effet, elle est aussi pertinente pour révolutionner les technologies de l'information que les matériaux et la médecine (Meyer & Persson, 1998).

<sup>2</sup> Une recension de sous-catégorisations de la nanotechnologie se trouve à l'annexe 11.

MOCVD), points quantiques, nanotubes, tubes quantiques, tiges et fibres, nanopolymères, couches de graphites, fullerènes, nanocristaux. Source: Porter et al. (2007)

Nous trouvons que cette méthode de classification est très bien adaptée à nos besoins, à cause de sa pertinence, sa simplicité et le fait que les catégories présentent un chevauchement minimal entre elles.

## **2.1.2 Biotechnologie**

### **2.1.2.1 Définition**

La biotechnologie est définie comme étant l'application de la science et de la technologie à des organismes vivants, ou à une de leur partie, produit ou modèle, afin de modifier des parties matérielles vivantes ou non pour la production de connaissances, biens et de services. (OECD, 2010).

### **2.1.2.2 État de l'art de la biotechnologie**

Considérant sa définition, la biotechnologie est évidemment plus vieille que la nanotechnologie et certains diront qu'elle prend ses racines dans des processus agricoles tels des programmes d'élevage qui datent de plusieurs milliers d'années. La biotechnologie a grandement évolué depuis ses humbles débuts et englobe désormais tout un éventail d'applications, e.g. la biotechnologie computationnelle aussi connue sous le nom de la bioinformatique, des applications marines, des processus médicaux, des applications industrielles en production chimique. La biotechnologie moderne, ou simplement la biotechnologie, tel que nous y ferons référence dans la suite de ce document, est née avec l'avènement de l'ADN recombinant en 1973 et est le fruit d'une collaboration entre deux Californiens, Stanley Cohen & Herbert Boyer (Darby & Zucker, 2003).

Les découvertes et percées clefs ont été incessantes depuis la fin des années '70 et le début des années '80 et ceci est non seulement dû au travail acharné des scientifiques impliqués mais aussi à l'important financement privé et public de la biotechnologie autant dans des milieux industriels qu'universitaires. En fait, les dépenses totales en R&D biotechnologique dans le secteur privé

canadien étaient de 944,5 million USD en 2008 (OECD, 2011)<sup>3</sup>. Par ailleurs, selon l'OCDE (2011), la R&D publique en biotechnologie représente près de 6,69 % du total des dépenses publiques en R&D. De plus, le Canada est l'un des meneurs en ce qui concerne l'indicateur de *l'avantage technologique relevé*<sup>4</sup> en biotechnologie. En termes d'application, à compter de 2006, 87,3% des investissements en R&D ont été dans le secteur de la santé, 9,2 % en agriculture et le reste est éparpillé à travers les ressources naturelles, l'environnement, les processus industriels, la bioinformatique et autres (OCDE, 2011).

### 2.1.2.3 Sous-domaines

La classification de brevets biotechnologiques peut se faire de plusieurs manières, soit par industrie, par définition, par application, etc.<sup>5</sup> Nous choisissons la classification par définition de techniques biotechnologiques de l'OCDE (2005). Les avantages de cette méthode sont que, de par sa nature, il n'y a pas de chevauchement entre les sous-catégories (qui sont d'ailleurs peu nombreuses). De plus, elle contient une catégorie commune à la classification de la nanotechnologie de la *GT CNS-ASU Group* : la nanobiotechnologie. Voici les catégories employées par l'OCDE (2005) définies comme suit :

- (1) ADN/ARN : Génomique, pharmacogénomique, sonde génétique, ingénierie génétique, séquençage/synthèse/amplification d'ADN/ARN, profilage d'expression génétique et utilisation de techniques antisens;
- (2) Protéines et autres molécules : séquençage/synthèse/ingénierie de protéines et de peptides (incluant des molécules d'hormones large), amélioration des méthodes de libération de médicaments à molécules larges, protéomiques, isolation et purification de protéines, signalisation, identification de récepteurs cellulaires;

---

<sup>3</sup> Selon le Directorate for Science Technology and Industry (DSTI) de l'OCDE.

<sup>4</sup> Le (revealed technological advantage indicator) est calculé comme la proportion des brevets biotechnologiques d'un pays relativement à proportion de la biotechnologie dans le total des brevets de [http://www.oecd.org/document/30/0,3746,en\\_2649\\_34537\\_40146462\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/30/0,3746,en_2649_34537_40146462_1_1_1_1,00.html)

<sup>5</sup> Les options de catégorisation par Statistique Canada, Derwent et l'OCDE se trouvent à l'annexe 12.

- (3) Culture et ingénierie cellulaire et tissulaire : Culture tissulaire et cellulaire, ingénierie tissulaire (incluant l'échafaudage tissulaire et ingénierie biomédicale), fusion cellulaire, stimulants vaccinaux/immunitaires, manipulation des embryons;
- (4) Techniques de processus biotechnologiques : Fermentation en utilisant des bioréacteurs, biotransformation, biolixiviation, biopulpage, biodésulfuration, biorestauration, biofiltrage, phytoremédiation;
- (5) Vecteurs de gènes et d'ARN : Thérapie génique, vecteur viral;
- (6) Bioinformatique : Construction de bases de données sur les génomes, séquences de protéines, modélisation de processus biologiques complexes, incluant la biologie des appareils anatomiques.
- (7) Nanobiotechnologie : Outils et processus de nano/microfabrication pour la construction de dispositifs pour étudier des biosystèmes et des applications dans la libération de médicaments, diagnostics, etc. Source : OCDE (2005)

### **2.1.3 Note sur la nanobiotechnologie**

Tout dépendant à qui nous demandons de définir la nanobiotechnologie, sa définition peut grandement varier. Et même, selon plusieurs, la bionanotechnologie n'est pas synonyme de nanobiotechnologie. Il y a notamment Gazit (2007) qui offre la distinction concise suivante. Il emploie le mot nanobiotechnologie pour décrire l'application de techniques nanotechnologiques pour le développement et l'amélioration de produits et processus biotechnologiques et il réserve l'utilisation du mot bionanotechnologie pour décrire l'utilisation de composantes de base biologiques et l'utilisation d'activités et de spécificités biologiques pour le développement à l'échelle nanométrique (Gazit, 2007).

À toutes fins pratiques, nous fusionnons les éléments regroupés sous chacune de ces définitions dans la même catégorie, que nous choisissons de nommer nanobiotechnologie. Ceci nous évite des complications potentielles liées une analyse de sémantique trop pointue.

## **2.2 Innovation**

Selon l'Organisation de Coopération et de Développement Économique (OCDE, 1997) « On entend par *innovation technologique de produit* la mise au point/commercialisation d'un produit

plus performant dans le but de fournir au consommateur des services objectivement nouveaux ou améliorés. Par *innovation technologique de procédé*, on entend la mise au point/adoption de méthodes de production ou de distribution nouvelles ou notablement améliorées. Elle peut faire intervenir des changements affectant – séparément ou simultanément – les matériels, les ressources humaines ou les méthodes de travail ». Ce sont ces deux types d'innovation qui nous intéressent dans le cadre de cette recherche.

## 2.2.1 Modèle d'innovation

Typiquement amener une innovation à terme implique deux types fondamentaux de transferts de technologies ou de connaissances. Ces derniers sont explicités par Butler (2011) dans la Figure 2-1 ci-dessous<sup>6</sup>.

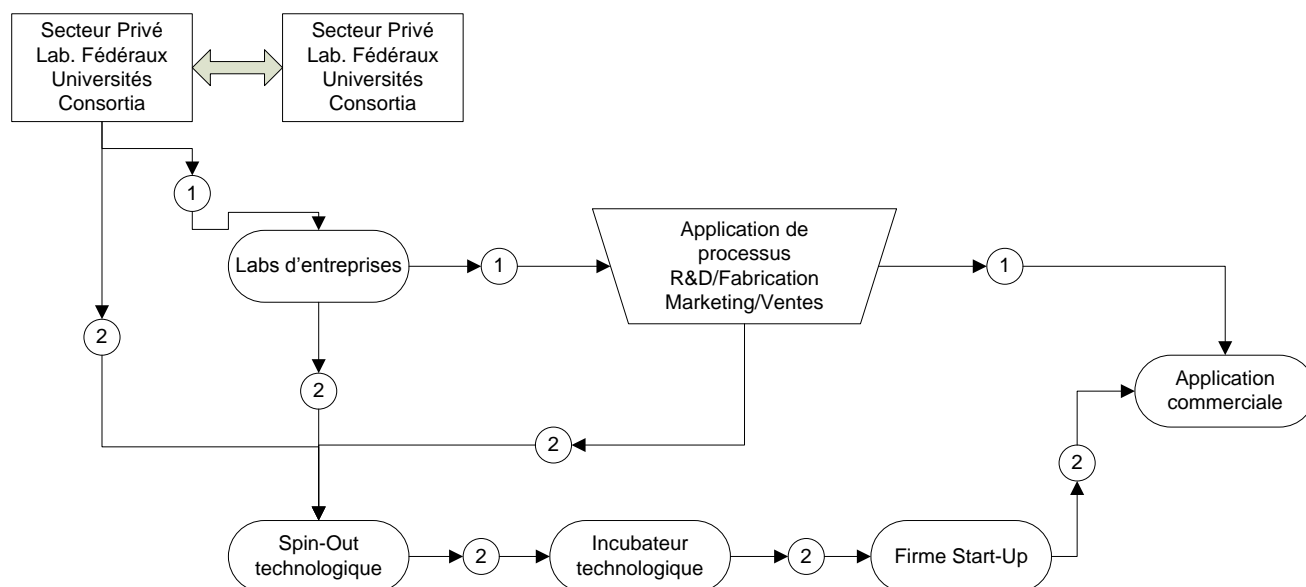


Figure 2-1: Deux types de base de transfert de connaissances et de technologie à la commercialisation (Butler, 2011)

<sup>6</sup> Dans ce modèle les connaissances sont produites par divers acteurs, soit en collaboration ou non, et prennent ensuite l'un de deux chemins principaux. (1) Elles sont appropriées par des laboratoires d'entreprises et exploitées d'une manière industrielle classique. (2) Les connaissances sont à la base d'une *Spin-Out* (ou *Spin-off*); elles sont ensuite développées de manière indépendante jusqu'à la mise en marché du produit final.



Cependant, lorsque vient le temps d'analyser l'innovation dans des domaines de haute technologie, tels la nanotechnologie ou la biotechnologie, le transfert des technologies et leurs extrants méritent une attention particulière, d'où l'intérêt du modèle de la triple hélice.

Il est reconnu que le modèle d'innovation typique a évolué d'un schéma linéaire établi dans les années 1960 pour inclure les interactions industrie-université dans les années 1990, mais aussi les réseaux scientifiques, la stratégie, l'analyse structurelle de son industrie et de compétiteurs (Van Looy *et al.*, 2004) et possède maintenant la forme d'une triple hélice. Ce concept de triple hélice marque la synergie entre le gouvernement, l'industrie et l'université dans le développement de l'innovation; les contributions de ces 3 entités étant également importantes pour la progression d'une société de connaissance.

Il y a trois éléments de base associés à cette vision :

- (1) Un rôle plus important pour l'université dans l'innovation, égale à la contribution de l'industrie et du gouvernement;
- (2) Un mouvement vers des relations de collaboration entre les 3 entités en question, dans lequel les politiques d'innovation sont davantage le fruit d'interactions plutôt que de prescription gouvernementales;
- (3) L'adoption de rôles non-traditionnels (habituellement réservés aux autres) par chacune des entités (Triple Helix, 2011).

Certains diront même qu'une quatrième hélice est à inclure dans le modèle, la société (Rieu, 2011). Cependant l'apport de la société étant difficilement quantifiable, elle ne sera pas traitée dans ce projet.

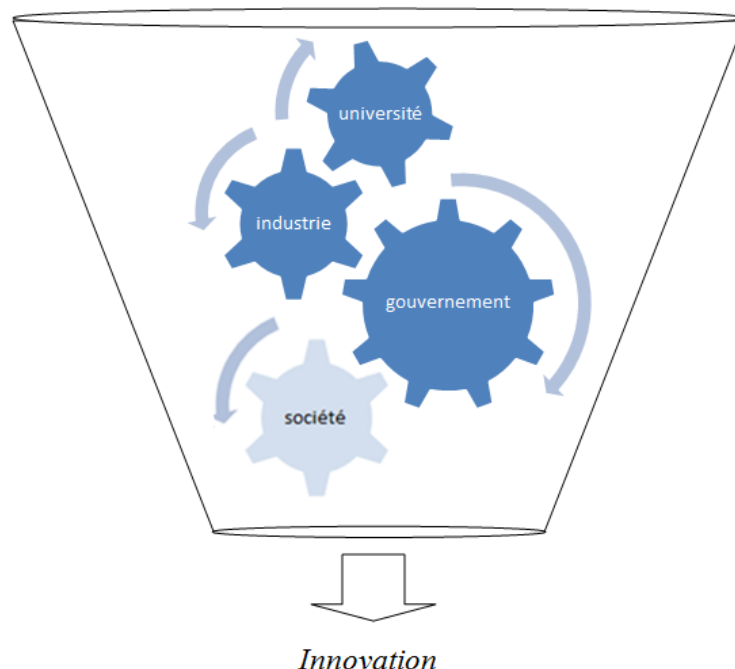


Figure 2-2: Modèle Triple (Quadruple) Hélice

Même si il y a eu une diversification concrète des lieux de production de connaissances, plusieurs dont Godin & Gingras (2000) trouvent que les universités maintiennent une position clef dans ce système de production et que les autres acteurs impliqués dépendent grandement de leur savoir-faire (Godin & Gingras, 2000). En fait, Godin & Gingras trouvent que contrairement à la croyance populaire, plutôt que de conduire à une marginalisation de l'un des acteurs impliqués dans le processus de création des connaissances, l'étude de l'évolution des relations entre gouvernements, industries et universités ne fait qu'identifier un renforcement des interactions entre ces-derniers. Ainsi, ceci vient renforcer le fait que le rôle joué par les chercheurs universitaires dans la contribution à l'innovation est non-négligeable et même très pertinente.<sup>7</sup>

Fishbine (2002) emprunte une vision un peu moins macro et énonce qu'il y a cinq acteurs principaux (ou programmes de R&D) qui ont une implication en innovation nanotechnologique. Ils sont spécifiques mais restent assez généraux pour tout de même être applicables en biotechnologie.

---

<sup>7</sup> Ce que nous pouvons donc retirer de ceci est que les différents acteurs impliqués en innovation ne travaillent pas en silo, isolés les uns des autres, mais sont plutôt dépendantes entre eux.

- (1) Le premier est le milieu universitaire, dominé par des individus qui travaillent dans un environnement académique et qui s'efforcent à résoudre les problèmes de nature fondamentale entourant une technologie émergente.
- (2) Le deuxième type est le système de laboratoire national, qui est souvent en concurrence avec le milieu académique pour les fonds gouvernementaux.
- (3) Le troisième acteur est le milieu industriel de recherche classique. Principalement composé de grandes entreprises, e.g. celle de la liste *Fortune 1000*.
- (4) La quatrième organisation impliquée dans le développement technologique est l'organisme à but non-lucratif (OBNL) qui tente de créer un climat de politique et d'investissement favorable pour le financement de nouvelles initiatives.
- (5) La cinquième et dernière catégorie contient les entrepreneurs dans qui plusieurs actionnaires investissent pour voir la création de nouveaux produits et ultimement un profit. (Fishbine, 2002)

Peu importe, les acteurs impliqués reviennent comme étant les mêmes et Roco (2001) a identifié plusieurs tendances stratégiques de ceux-ci pour l'investissement en R&D nanotechnologique. Cette stratégie permet le développement du plein potentiel de la technologie et est axée sur diverses lignes directrices:

- (1) différentes grappes de R&D selon le pays;
- (2) la formation du personnel est clef au succès à long terme; Cette formation a lieu tant en entreprise que dans le milieu universitaire.
- (3) certains obstacles techniques et scientifiques communs englobent de grands objectifs humains; d'où le besoin d'agréger recherche fondamentale et appliquée.
- (4) une concentration sur la nanoproduction;
- (5) des partenariats afin d'encourager l'interdisciplinarité et l'intégration des activités de recherche;
- (6) la collaboration internationale.

Maintenant que nous avons une idée globale des acteurs engagés et de leurs rôles respectifs en innovation nous allons explorer les mécanismes de financement.

### 2.2.2 Financement de chercheurs universitaires et innovation

Les scientifiques universitaires reçoivent des fonds, sous la forme de subventions ou de contrats, de plusieurs types d'organisations et pour plusieurs raisons. Les universités financent les universitaires car ils ont un intérêt direct dans le succès de leurs chercheurs. Sans financement approprié, la plupart des universitaires ne serait pas aussi accomplis dans leur carrière respective. Ceci est tout particulièrement vrai dans les cas de la biotechnologie et de la nanotechnologie où l'infrastructure et l'équipement nécessaire sont onéreux (Robinson *et al.*, 2007). De plus, un nombre accru de publications, rendu possible par l'accès à des fonds recherche, reflète très bien sur l'institution universitaire à laquelle le chercheur est affilié. Ceci étant dit, les universités ne sont pas les seuls enthousiastes face à l'idée de financer des activités de recherche d'universitaires. Les firmes, fondations, associations, sociétés, établissements de santé et sociétés d'état sont des exemples d'autres organisations désireuses de pourvoir des fonds.

En particulier, Bonaccorsi & Piccagula (1994) énoncent que les firmes ont plusieurs raisons de vouloir collaborer avec des laboratoires académiques et que due à la croissance de l'écart entre les coûts de recherche et le financement gouvernemental, les universitaires ont même commencé à chercher des fonds auprès des firmes privées pour financer leur recherche. Ils continuent en expliquant par contre que les motivations d'une firme à entamer une relation avec une université ne se limitent pas là. Entre autres, les firmes requièrent la construction et l'amélioration de connaissances scientifiques afin de reconnaître et d'exploiter diverses opportunités technologiques. Il ne faut pas négliger non plus le fait que plusieurs percées scientifiques non-triviales ont été le produit d'une résolution de problèmes techniques communs auxquels les firmes faisaient face soit dans leurs activités de conception ou de production. Il a d'ailleurs été suggéré par Bonaccorsi & Piccaluga (1994) que des liens externes avec les universités ne sont pas des remplacements du financement de recherche fondamentale interne, entre autre parce que la capacité d'absorption de la firme est dépendante de ses investissements précédents en R&D. Par contre de manière générale, George *et al.* (2002) trouvent que les firmes avec des liens universitaires ont des coûts de R&D plus faibles et de plus haut niveaux de production d'innovation. Concernant la collaboration de firmes et d'universités, Meyer (2006) souligne que :

- (1) la croissance dans la spécialisation de la connaissance pousse les firmes à accroître leur dépendance sur une combinaison de R&D interne et externe (Brusoni *et al.*, 2001; Grandstrand *et al.*, 1997; Langlois, 1992)
- (2) il y a un genre de couplage dans le sens où les firmes maintiennent des amitiés avec des sources indépendantes externes, e.g. des fournisseurs, universités, etc. qui lui permettent de détecter des changements technologiques dans des secteurs d'activités non-limités aux leurs.

Une autre raison, qui justifie les relations de firmes avec les universités est le fait que des applications importantes sont souvent le résultat de recherche effectuée simplement par curiosité et que ce type de recherche est le plus souvent et naturellement externalisée aux universités (Pavitt, 1990; Bonaccorsi & Piccaluga, 1994).

En outre, le financement privé de chercheurs universitaires est particulièrement important en nanotechnologie, comme en biotechnologie; Narin *et al.* (1997) montrent que la connexion intra-nation entre science fondamentale (le plus souvent réalisée en milieu universitaire) et la science appliquée (plus souvent réalisé en entreprise) est significative dans des domaines de haute technologie, rendant ainsi les liens université-industrie particulièrement pertinents dans ces domaines (Meyer & Persson, 1998). De plus, la recherche exhaustive effectuée par Zucker *et al.* (1998) et Murray (2004) a démontré que la performance de compagnies dans le domaine de la biotechnologie est augmentée par l'existence de liens avec des scientifiques universitaires (Casper, 2007). En ce qui concerne les nanotechnologies, elles sont jusqu'à un certain point davantage multidisciplinaires que d'autres domaines et requièrent de l'équipement onéreux que les acteurs impliqués ont tout intérêt à partager (Robinson *et al.*, 2007).

Pour ce qui est des universitaires en tant que tels, ils acceptent des fonds de firmes pour plusieurs raisons. Premièrement, Gulbrandsen & Smeby (2002) suggèrent que contrairement à la croyance populaire, les relations université-industrie ne nuisent pas aux objectifs, ni aux récompenses académiques. De plus, Gulbrandsen & Smeby (2005) trouvent que les universitaires qui bénéficient de financement industriel externe publient plus d'articles que leur pairs, confirmant ainsi des résultats obtenus par Godin (1998) et Geuna & Nesta (2006). Il est important de mentionner cependant que les scientifiques ont davantage tendance à vouloir collaborer avec des

firmes après avoir remarqué que leurs collègues ont plus de publications et de citations ce faisant (Darby & Zucker, 2003).

Voyant comment le sujet du financement universitaire devient progressivement significatif, nous aimerions examiner si ce financement affecte vraiment de manière positive l'innovation et la réussite pour tous les acteurs impliqués. Ceci nous amène à notre première question : jusqu'à quel point le financement d'un chercheur universitaire mène-t-il au brevetage?

### **2.2.3 Réseaux**

Zucker & Darby (1996) proposent que les découvertes scientifiques ne sont pas des morceaux d'informations ambulantes qui, une fois découvertes, se propagent comme une contagion anonymement et pour laquelle les identités des personnes impliquées ne sont pas importantes. Tout au contraire. Les réseaux chercheurs-chercheurs, chercheurs-organisations et organisations-organisations sont d'une importance immense dans le processus d'innovation dans les domaines de haute technologie tels la nanotechnologie et la biotechnologie. En outre, des recherches considérables ont démontré que la performance de compagnies œuvrant dans le domaine de la biotechnologie est améliorée par l'existence de liens avec des scientifiques académiques (Zucker et al., 1998; Murray, 2004). La R&D collaborative entre chercheurs et entre firmes permet entre autres :

- (1) l'accélération de la commercialisation de nouvelles technologies;
- (2) la facilitation et l'accélération du transfert de résultats de recherche des universités ou des laboratoires publics à l'industrie;
- (3) de supporter l'accès des firmes industrielles aux capacités R&D d'installations de recherche nationales;
- (4) de supporter la création d'une « vision » technologique commune dans une industrie, ce qui peut guider la R&D et les investissements publics et privés reliés (Mowery, 1998).

Ainsi la proximité sociale combinée avec la proximité géographique engendrent des transferts de connaissances plus efficaces (Owen-Smith and Powell, 2004; Sorensen and Stuart, 2001)

D'ailleurs Breschi et Catalini (2010) trouvent que les scientifiques-inventeurs (en moyenne) sont plus intermédiaires et sont positionnés plus centralement par rapport à leur collègues qui ne sont

qu'auteur ou inventeur dans les réseaux de co-publication comme dans les réseaux de co-invention<sup>8</sup>. Les auteurs concluent que même s'ils ont de différents objectifs et structures incitatives les deux communautés de chercheurs (science ouverte et industrie) sont connectées à un degré relativement élevé. Ainsi, étant donné l'importance des réseaux sociaux dans le développement des domaines à l'étude, les réseaux scientifiques seront couverts dans ce mémoire (les définitions des mesures de réseaux se trouvent à la section 3.4.1.2.)

## **2.2.4 Fondement de la comparaison de la nanotechnologie et de la biotechnologie**

Il a été mentionné par Darby & Zucker (2003) que parce que les découvertes scientifiques qui sous-tendent la nanotechnologie et la biotechnologie sont essentiellement des méthodes d'invention, c'est-à-dire qu'ils créent des opportunités technologiques mais aussi l'appropriabilité à travers plusieurs disciplines, ils s'attendent à observer un développement similaire. À bien des égards, la croissance de la nanotechnologie est considérée similaire à celle de la biotechnologie (Rothaermel & Thursby, 2007). Darby et Zucker ont effectué des analyses de brevetage, de publication et d'entrée de start-ups à proximité d'universités et mis en évidence plusieurs tendances similaires dans le domaines de la biotechnologie (Darby & Zucker, 2003; Zucker *et al.*, 1998; Rothaermel et Thursby, 2007). Certains pourraient alors faire valoir que si tous ces éléments étaient similaires alors le processus de financement et brevetage subséquent devrait l'être aussi.

Dans ce travail, nous utilisons le brevetage comme indicateur principal d'innovation. Nous le faisons pour les raisons suivantes : premièrement comme le soulignent Rothaermel et Thursby (2007), le brevet représente une instance où il y a une invention nouvelle, non-triviale et utile; deuxièmement, les brevets contiennent autant de l'information sur les inventeurs (i.e. les scientifiques) et les cessionnaires (i.e. firmes et autres organisations) et représentent essentiellement la seule façon de lier les acteurs entre eux tout en modélisant l'interaction

---

<sup>8</sup> En d'autres mots, ils sont mieux placés pour générer des connaissances novatrices que leurs collègues isolés soit au réseau d'invention soit au réseau de publication.

industrie-université. En fait, Darby et Zucker (2003) mentionnent que les brevets sont une arme souvent utilisée dans la protection de produits biotechnologiques. Ceci est vrai surtout après l'invention cependant car *l'exclusivité naturelle*<sup>9</sup> fournit une protection initiale en ralentissant la diffusion de nouvelles connaissances (Zucker et Darby, 1996; Zucker, Darby et Brewer, 1998; Darby et Zucker, 2003) en nanotechnologie comme en biotechnologie. Darby et Zucker (2003) mentionnent qu'étant donné un état similaire de maturité, et prenant en compte le décalage temporel, la croissance scientifique et le brevetage de la nanotechnologie suit celle de la biotechnologie. De plus, en utilisant les brevets comme mesure d'extrait innovant, Rothaermel et Thursby (2007) illustrent les séries temporelles de la nanotechnologie et de la biotechnologie et montrent que la nanotechnologie est décalée de 4 ans, et qu'il y aurait un coefficient de corrélation  $r=0,95$  avec la série de biotechnologie. Cette dernière déclaration fournit un autre facteur principal supportant l'hypothèse que le développement de la nanotechnologie suit de très près celui de la biotechnologie, du moins en ce qui concerne les brevets.

Ceci nous amène à la deuxième question à laquelle nous aimerions répondre : est-ce que le processus d'innovation comprenant le financement-brevetage en nanotechnologie diffère de celui en biotechnologie?

## 2.3 Synthèse

Dans cette section nous avons fait le point sur les définitions des domaines à l'étude, leur importance tant au niveau technologique qu'économique et les facteurs qui rendent leur développement si similaire. Nous avons aussi pu définir le processus d'innovation en nanotechnologie comme en biotechnologie, en soulignant l'importance de certains éléments, dont le rôle des chercheurs universitaires et des réseaux de collaboration. Nous avons donc une base sur laquelle nous pouvons construire nos hypothèses, qui serviront de guide pour la suite de ce mémoire.

---

<sup>9</sup> Ceci est issu du fait qu'avant la codification, les découvertes notables impliquent souvent des connaissances tacites importantes qui sont incarnées d'abord dans les découvreurs et qui se font transmettre par *learning by doing* (Darby & Zucker, 2003).



## CHAPITRE 3    DONNÉES, HYPOTHÈSES ET MÉTHODOLOGIE

Nous commençons d'abord par une évaluation sommaire de la faisabilité du projet de recherche désiré. Un des avantages de ce projet de recherche vient du fait que la plupart des données de financement nécessaires ont déjà été saisies par Larivière et reformatés par Beaudry et Clerk-Lamalice et celles de brevetage ont été extraites par Barirani. Les données adéquates étant à notre disposition, nous avons commencé par faire la mise au point de l'ensemble des données et aussi la compilation de statistiques et graphiques descriptifs (Annexe 1 et Annexe 2) auxquelles nous ferons référence dans la section 3.3 de ce mémoire.

### 3.1 Hypothèses

Les similarités entre la nanotechnologie et la biotechnologie mises en évidence plus haut dans la littérature, ne sont pas négligeables. Cependant, il existe aussi des justifications pour suggérer qu'ils ne sont pas aussi similaires que l'on puisse le croire. Rothaermel & Thursby (2007) trouvent que le brevetage en biotechnologie est expliqué par des connaissances acquises à l'extérieur de la firme, spécialement à travers des alliances de R&D, ainsi que de leur base de connaissances antérieures. En divisant une période de 20 ans en deux, 1980-1990 et 1990-2000, ils trouvent que la biotechnologie était plus enclin à des alliances externes de R&D dans la première période et plus enclin à faire de la R&D interne dans la deuxième période. La nanotechnologie est différente dans le sens où elle était plus encline à dépenser pour effectuer de la R&D à l'interne dans les deux périodes<sup>10</sup>. Cela suggère que les entreprises de biotechnologie se sont appuyées sur des alliances et des acquisitions beaucoup plus longtemps que les entreprises en place dans les nanotechnologies. Cela nous conduit à proposer l'hypothèse suivante:

---

<sup>10</sup> Ceci s'explique par le fait que le temps avant que l'instrumentation pour la première technologie permettant la biotechnologie, i.e. le séquençage de gène automatique, était disponible pendant 20 ans (Zucker *et al.*, 1998) et que la technologie permettant la nanotechnologie, i.e. le microscope à effet tunnel, développé en 1981 et l'instrument employé, le microscope à effet atomique a été commercialisé en 1989 (Darby et Zucker, 2003). Ainsi Rothaermel et Thursby (2007) énoncent que les entreprises établies en biotechnologie devraient s'appuyer sur des alliances, des fusions et des acquisitions beaucoup plus longtemps que des firmes en nanotechnologie.

**Hypothèse 1 :** Même si elles présentent plusieurs similarités, la nanotechnologie et la biotechnologie suivent un développement d'innovation différent.

La nature de notre analyse est relativement exploratoire. Cependant, il ne serait pas exagéré de croire que parce que la biotechnologie et la nanobiotechnologie sont de nature similaire, ils possèdent le même schéma d'innovation. Nous aurons des données limitées à ce sujet cependant, car comme le mentionne Hullman (2007), même s'il y a eu un développement de plus en plus dynamique, le Canada a tendance à se spécialiser en nanomatériaux et nanodispositifs. Quoi qu'il en soit, nous proposons l'hypothèse suivante:

**Hypothèse 2 :** La nanobiotechnologie suit un schéma d'innovation similaire à celui de la biotechnologie.

Afin d'évaluer l'importance du financement pour la production scientifique nous créons un concept que nous nommons *boucle d'innovation*. Essentiellement, une boucle d'innovation est une instance dans laquelle un chercheur universitaire participe au brevetage d'une invention qui appartient à l'organisation qui a initialement financé le chercheur en question. Ce concept est illustré dans le graphique ci-dessous.

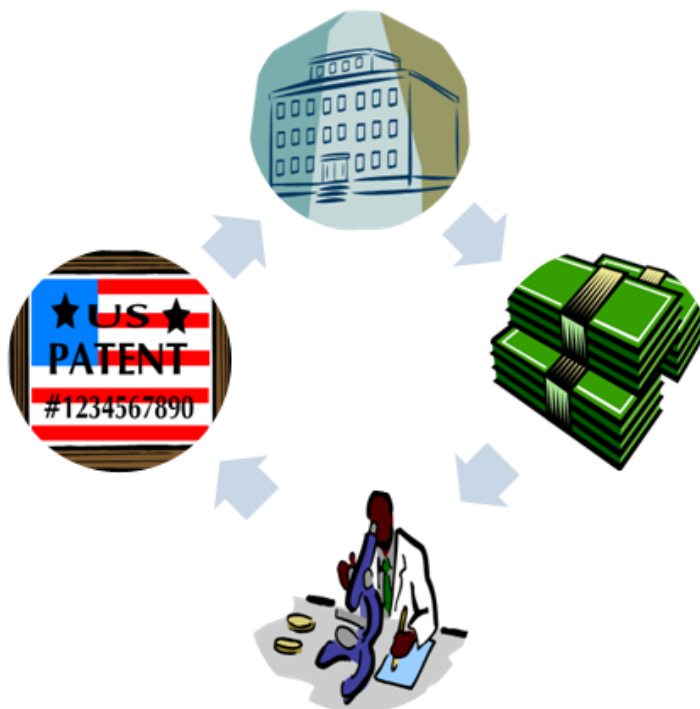


Figure 3-1 : Boucle d'innovation

Nous reconnaissons que ceci ne représente qu'une partie du processus d'innovation et qu'elle néglige certains aspects importants tels, la collaboration impliquant les transferts de connaissances tacites, etc. mais elle est tout de même amplement adéquate pour effectuer une comparaison approfondie des domaines à l'étude.

Nous pouvons considérer que puisque les brevets qui font partie de boucles et dont les cessionnaires sont des firmes sont des produits de collaboration, ils devraient présenter une meilleure qualité que les brevets qui sont hors boucle.

**Hypothèse 3 :** Autant les brevets en nanotechnologie qu'en biotechnologie générés au sein de boucles d'innovation sont de plus grande qualité que ceux produits hors boucle, i.e. (a) ils ont plus de revendications et (b) un plus haut taux de citations.

Comme il y a un intervalle entre le financement du chercheur et le brevetage qui en découle, nous aimerions tester si oui ou non l'étendue de cet intervalle implique une différence dans le degré d'application des connaissances qui y sont associées. Ceci est une hypothèse plutôt empirique que nous trouvons néanmoins pertinente pour comparer la nanotechnologie et la biotechnologie. De plus, nous pouvons contribuer à l'analyse de similarité entre ces deux domaines en analysant la différence du degré d'application des recherches entre eux. Essentiellement, parce que leurs applications couvrent différents secteurs, certains, ayant recours à des connaissances plus appliquées que d'autres, les articles publiés de l'un devrait être d'un différent degré d'application que l'autre.

**Hypothèse 4 (a)** Les inventions ayant résulté d'un plus long intervalle de temps entre le financement et la demande d'application du brevet sont d'une nature plus appliquée que celles dont l'intervalle de temps était plus court. (b) Les connaissances associées aux brevets en nanotechnologie sont d'une nature plus appliquée que celles associées aux brevets en biotechnologie.

Il est reconnu que divers facteurs, incluant le financement d'un chercheur, sa position dans le réseau, ses affiliations universitaires et d'autres caractéristiques personnelles, expliquent sa capacité à produire des extraits scientifiques. Nous voulons revenir sur certains de ces facteurs pour déterminer la qualité de ces extraits ou plus particulièrement des brevets. En nous basant sur la littérature et les résultats précédents, nous proposons l'hypothèse suivante portant sur ce sujet.

**Hypothèse 5** Des différents facteurs servent à expliquer le nombre de revendications d'un brevet et le nombre de citations qu'il reçoit.

## 3.2 Objectifs

### Objectif 1

Nous voulons tout d'abord comparer la similarité scientométrique et le développement des domaines et sous-domaines de la nanotechnologie avec celles de la biotechnologie. L'objectif 1 se divise en deux catégories : (a) les statistiques comparatives et (b) la modélisation économétrique.

- (a) Ceci nécessite de comparer le degré d'application des connaissances, les intervalles de temps entre le dépôt et l'octroi des brevets, le nombre de chercheurs par brevet, les moyennes d'investissement financier par chercheur, de comparer les mesures de qualité entre les domaines (les nombre de revendications, les nombres de citations), et les moyennes des degrés d'application. Pour approfondir le sujet, nous avons comme objectif de comparer les mesures de qualité des brevets issus des boucles d'innovation et de comparer le degré d'application des connaissances associées aux brevets issus des boucles d'innovation par rapport aux autres brevets.
- (b) Nous désirons ensuite modéliser les facteurs influençant la production de brevets par les chercheurs universitaires, tout en considérant leur qualité. Nous cherchons ainsi à cibler les éléments qui contribuent à la production d'innovation d'un chercheur et vérifier s'ils sont similaires d'un domaine à l'autre.

Cette première analyse nous permet de trouver les similarités et différences associées au processus d'innovation en nanotechnologie et en biotechnologie, mais aussi de nous apercevoir qu'il y a une divergence des résultats associés à deux mesures de qualité reconnues. C'est pour cette raison que nous abordons une comparaison complète de ces mesures de qualité plus loin.

### Objectif 2

Tel que mentionné, le deuxième objectif porte sur les mesures de qualité. Cette étude enrichit la littérature relative à la qualité des brevets et nous permet d'approfondir l'analyse d'innovation en

nanotechnologie et en biotechnologie. Nous voulons identifier les facteurs affectant chacune des mesures de qualité et évaluer leur impact.

### **3.3 Données**

Pour mener nos analyses, nous employons des données provenant de sources variées. Dans les pages qui suivent nous ferons un survol de ces données et des métadonnées pertinentes qui y sont liées.

#### **3.3.1 Financement**

Les données de financement ont été extraites du Système d'Information sur la Recherche Universitaire (SIRU) qui est le système responsable de la compilation du financement gouvernemental, industriel et universitaire (sous la forme de subventions et de contrats) octroyé aux chercheurs universitaires québécois. SIRU est géré par le Ministère de l'Éducation, du Loisir et du Sport (MELS) et fournit les détails concernant la collaboration financière entre chercheurs universitaires et des personnes morales publiques et privées. En effet, tous les contrats et subventions qui transitent par les comptes de l'université pour financer la recherche universitaire sont comptabilisés par ce système. Pour les fins de notre analyse, nous utilisons les données contenues dans l'intervalle de temps entre 1983 et 2005. Les données passées 2005 ne sont pas nécessaires puisqu'un intervalle de 22 ans est amplement suffisant pour observer des tendances et cartographier l'évolution du financement. La table SIRU (voir annexe 1) contient chaque instance pour laquelle un chercheur universitaire québécois a reçu du financement en nanotechnologie ou biotechnologie. Les montants de financement varient entre Pour chaque instance, nous avons les informations suivantes :

- (1) L'année de financement ;
- (2) Le prénom et nom de famille du scientifique recevant les fonds ;
- (3) Le titre du projet sur lequel le chercheur travaillait;
- (4) Le nom de l'institution d'affiliation du chercheur;
- (5) Le nom du département d'affiliation du chercheur;

- (6) Le nom de l'organisation pourvoyeur de fonds;
- (7) La catégorie/secteur de l'organisation pourvoyeur;
- (8) L'origine de l'organisation pourvoyeur;
- (9) Le type de financement octroyé.

### 3.3.2 Brevetage

Les données de brevetage ont été extraites de la *United States Patent & Trademark Organization* (USPTO) à l'aide du logiciel PatentBot développé par Barirani (2008) qui extrait les brevets contenant les mots-clefs compilés et utilisés en nanotechnologie par Mogoutov et Kahane (2007), Porter et al. (2008), Schmoch et al. (2003), Alencar et al. (2007), Zitt & Bassecoulard (2006) & OECD (2009) et la définition de l'OCDE en biotechnologie. Il faut mentionner que l'Office de la Propriété Intellectuelle du Canada (OPIC) était un choix potentiel intéressant pour obtenir les données de brevetage, surtout considérant que notre étude porte sur des inventeurs canadiens (québécois). Par contre, l'OPIC ne rend pas autant de données disponibles que l'USPTO, e.g. les adresses des inventeurs sont manquantes rendant l'association des chercheurs-inventeurs plus difficile et moins rigoureuse. De plus, comme il est mentionné par Schiffauerova & Beaudry (2009), l'USPTO est un substitut acceptable et ce pour plusieurs raisons, dont le fait d'avoir un marché dix fois la taille de celle du Canada pour n'en nommer qu'une. 12 836 brevets en biotechnologie et 5 934 brevets en nanotechnologie sont extraits, couvrant un intervalle de temps entre 1976 et 2005 et les citations reçues jusqu'en 2009 sont gardées. Pour chaque instance de brevetage, nous possédons les données suivantes :

- (1) Le numéro d'identification du brevet;
- (2) Le titre du brevet;
- (3) Le résumé du brevet;
- (4) L'année d'application du brevet;
- (5) L'année d'octroi du brevet;
- (6) Le nombre de revendications associées au brevet;

- (7) Le(s) nom(s) du(des) cessionnaire(s) du brevet;
- (8) Le(s) nom(s) du(des) inventeur(s) du brevet;
- (9) Le type de brevet;
- (10) Le nombre de citations reçus par le brevet.

### 3.3.3 Publication

Les données de publication proviennent de SCOPUS, qui est une base de données régie par Elsevier contenant les résumés et les citations de plus de 18 000 journaux scientifiques. Ces publications sont extraites selon des mots-clefs spécifiques, comme pour le brevetage. En sélectionnant seulement les articles scientifiques pour lesquels au moins un auteur a une affiliation canadienne, nous obtenons 138 136 articles en biotechnologie et 42 543 articles en nanotechnologie entre 1929 et 2009. Les informations pertinentes retirées sont les suivantes :

- (1) Le titre de l'article scientifique;
- (2) Le nom de journal qui contient l'article;
- (3) La référence de l'article;
- (4) L'année de publication de l'article;
- (5) Les mots-clefs de l'article;
- (6) Le nom de l'auteur;
- (7) Le département d'affiliation de l'auteur;
- (8) Les citations reçues par chaque article selon l'année suivant la publication.

La fusion de ces trois bases de données en utilisant le scientifique comme point commun a requis de considérables ressources et plus de 200 heures de travail manuel (cf. section 3.4.2 pour la méthode d'appariement des scientifiques). Dans l'analyse économétrique nous limitons l'étude à quelques données qui répondent à des critères spécifiques. Premièrement, nous gardons seulement les chercheurs au Québec qui ont plus de 5 articles dans leur domaine, ceci est fait pour limiter la possibilité d'avoir un chercheur qui écrit hors de son domaine (mais dans un secteur connexe) dans l'échantillon. L'échantillon final contient 1460 scientifiques de 1996 à 2005 avec en moyenne 6 années de présence, et est donc composée de 9086 observations.

## 3.4 Méthodologie

Dans la première phase, nous compilons les mesures de réseaux et autres indices. Nous effectuons les tâches nécessaires à la création des boucles à l'aide de techniques et de méthodes éprouvées, telles celles employées par Calero et al. (2006), Daim et al. (2007), Hamilton (2003), Murray (2002) etc. Nous faisons ensuite un triage des brevets selon leur domaine d'application. Les analyses statistiques viennent à la fin. La majorité des données ont été compilées dans un serveur MySQL (Annexe 3) et traitées dans le logiciel MySQL Workbench 5.2.26 CE Révision 6432 et quelques traitements complémentaires ou explicatifs ont été effectués à l'aide de Microsoft Excel 2007. Cependant, toutes les analyses statistiques ont été effectuées dans le logiciel STATA IC v10 et/ou SE v11.

### 3.4.1 Indices scientométriques et mesures

Dans ce mémoire nous employons des méthodes bibliométriques, i.e. des méthodes qui englobent des techniques mathématiques et statistiques pour l'étude de publications et de communications dans la distribution de l'information (Diodato, 1994) et plus généralement des méthodes scientométriques, i.e. des méthodes de mesure et d'analyse de la science. Ces méthodes permettent d'évaluer l'avancement de la science et de la technologie tant à l'échelle du domaine de recherche qu'à celle de l'individu qui y œuvre. Les méthodes bibliométriques ont des limitations certes, mais elles sont utiles pour soulever des tendances ou pour quantifier certaines caractéristiques des données étudiées. Les éléments de la bibliométrie qui nous intéressent le plus ici sont les indicateurs de production (e.g. nombre de brevets), les indicateurs d'impact (e.g. le nombre de citations) et les réseaux de citations. Ci-dessous, dans les sections 3.4.1.1 et 3.4.1.2, se trouve l'information relative aux brevets et aux réseaux de citations. De plus, l'explication du degré d'application des connaissances associées aux articles scientifiques se trouve à la section 3.4.5.

#### 3.4.1.1 Brevets

Il est pertinent d'utiliser les brevets et les données associées aux brevets dans l'analyse de l'innovation car il est reconnu que jusqu'à 80% de toutes les connaissances technologiques sont contenues sous une forme ou une autre dans les brevets comme tels (Blackman, 1995) et elles



constituent une base solide pour l'étude et le suivi de la technologie (Teichert & Mittermayer, 2002). Les brevets sont donc des mines d'information lorsque proprement exploitées.

La qualité associée à un brevet peut être mesurée de plusieurs manières. En effet, les citations, le fait qu'un brevet soit triadique<sup>11</sup> ou non, le paiement associé à la protection du brevet et les revenus associés au brevet (Hall et al., 2001; Mowery et Ziedonis, 2002; Larivière, 2009) sont tous des moyens reconnus pour évaluer l'importance des brevets. Cependant, quelques limites s'imposent sur la plupart de ces instruments de mesure et des enquêtes approfondies seraient nécessaires pour recueillir l'information nécessaire.

Nous nous limitons à utiliser les revendications associées au brevet comme premier indicateur de qualité et les citations comme deuxième indicateur. Ces mesures ont été utiles pour mesurer la qualité par plusieurs dont Trajtenberg (1990), Lanjouw & Schankerman (2004) et Bonaccorsi & Thoma (2007).

Les citations de brevets peuvent être utilisées à plusieurs fins : e.g. Jaffe, Trajtenberg et Henderson, les utilisent pour analyser les spillovers de connaissances (Jaffe et al., 1993), elles peuvent être utilisées pour mesurer l'étendue des connaissances et même cartographier les réseaux de recherche d'un domaine spécifique. Dans notre cas, ce n'est pas le contenu des citations qui nous intéresse comme tel, mais le nombre de citations. Il est reconnu qu'un brevet recevant plus de citations a une plus grande importance, i.e. sa capacité à faciliter d'autres innovations est mise en valeur, ainsi plus grande est sa qualité.

Youtie et al. (2008) utilisent une pratique qui consiste à diviser le nombre de citations que reçoit un article par le nombre d'années dont il dispose pour recevoir des citations et ce pour limiter le biais du fait qu'un vieil article reçoit plus de citations qu'un nouvel article. Nous n'empruntons pas cette méthode, cependant il faut limiter le biais que pourrait engendrer un intervalle de temps non constant pour enregistrer le nombre de citations que reçoit un brevet<sup>12</sup>. Pour ce faire, nous

---

<sup>11</sup> Déposé aux États-Unis (USPTO), dans l'Union Européenne (EPO) et au Japon (JPO).

<sup>12</sup> Considérer le montant total de citations reçues par un brevet ne serait pas sage, car les vieux brevets auraient évidemment plus de citations que les brevets récents même s'ils sont de qualité inférieure.

gardons un intervalle constant de 3, 5 et 7 ans suivant l'octroi du brevet pour compter les citations.

Le nombre de citations employé ici est le nombre de citations effectuées par les brevets enregistrés à l'USPTO. Il existe plusieurs types de documents pouvant citer un brevet de la USPTO, tels que des brevets déposés ailleurs, des articles scientifiques etc., mais ils ne sont pas inclus dans cette analyse. Cette perte est minime par contre, car Meyer (2001) mentionne que même si la nano-science et la nanotechnologie semblent être relativement bien connectées par rapport à d'autres domaines, il n'y a qu'un très petit nombre de citations qui connectent les nano-brevets et les articles portant sur la nanoscience.

Comme pour les citations, ce n'est pas le contenu des revendications qui nous intéresse mais le nombre comme tel. Cette mesure a été utilisée par Schiffauerova & Beaudry (2009) et leur a permis d'effectuer certaines inférences, dont le fait que la qualité des brevets dont les propriétaires demeurent en dehors des grappes canadiennes est supérieure à la qualité des autres brevets.

Les revendications sont un bon indicateur de qualité puisqu'ils représentent l'étendue d'un brevet ; plus l'étendue est vaste plus le brevet est de qualité<sup>13</sup> (Tong & Frame, 1994 ; Zeebroeck & van Pottelsberghe de la Potterie, 2006).

### **3.4.1.2 Mesures réseaux**

Étant donné l'importance des effets de réseaux dans l'analyse de l'innovation en nanotechnologie et en biotechnologie (Schiffauerova & Beaudry, 2009), nous incluons certaines mesures de réseaux pour enrichir nos modèles économétriques (qui seront décrits plus loin). En vue du fait que notre étude porte sur les chercheurs universitaires, les données de réseaux utilisées sont celles compilées avec les publications d'articles scientifiques. Ce réseau permet d'avoir une meilleure

---

<sup>13</sup> Le seul désavantage d'employer le nombre de revendications comme mesure, est que nous ne considérons pas la valeur intrinsèque de chacune. En effet, les revendications n'ont pas toutes le même poids : certaines peuvent donc être de nature plus vaste ou utile que d'autres et donc dire de manière unilatérale qu'un brevet ayant 50 revendications est significativement de meilleure qualité qu'un brevet en ayant 49 ne serait pas nécessairement valide. Cela reste cependant une des meilleures mesures à notre disposition.

vue d'ensemble que le réseau de brevetage, afin d'obtenir des indices significatifs de transmission de connaissances entre chercheurs universitaires.

Plusieurs mesures ont été compilées par Beaudry & Clerk-Lamallice (2009)<sup>14</sup>, mais nous n'en utilisons que deux principales, soient : (1) le degré d'intermédiarité et (2) la densité égocentrique.

Le degré d'intermédiarité (*Betweenness centrality*) d'un chercheur est la proportion de toutes les géodésiques<sup>15</sup> entre les paires d'autres chercheurs qui incluent ce chercheur (De Nooy et al, 2005). L'idée sous-jacente est que plus un chercheur est central à son réseau plus il a une influence significative sur ce dernier (Freeman, 1979). C'est donc une mesure pertinente qui contribue à la représentation de l'importance sociale d'un chercheur dans un réseau. L'équation du degré d'intermédiarité est la suivante :

$$C_B(n_i) = \sum_{j < k} g_{jk}(n_i) / g_{jk}$$

Où

$g_{jk}(n_i)$  est le nombre de géodésiques qui lient deux chercheurs incluant le chercheur  $i$

$g_{jk}$  est le nombre de géodésiques qui lient deux chercheurs

La densité égocentrique (*Cliquishness*) est la densité du réseau entourant un chercheur (De Nooy et al, 2005). En d'autres mots, c'est la mesure du fait d'avoir un entourage connecté. Un chercheur dont tous les « voisins » sont connectés entre eux possède une densité égocentrique maximale (de 1). À toutes fins pratiques, plus la densité égocentrique d'un chercheur est grande, plus son réseau se trouve à être « intégré » et plus son importance sociale se trouve à être diluée, mais plus il est reconnu que la diffusion de connaissances dans ce réseau est facilitée.

---

<sup>14</sup> Incluant le degré de centralisation, la centralité d'intermédiarité, la centralisation d'intermédiarité.

<sup>15</sup> Une géodésique est le plus court chemin entre deux chercheurs.

### 3.4.2 Identification et assignation des noms de chercheurs

Comment faire correspondre un scientifique ayant produit une invention et qui est nommé comme inventeur ou co-inventeur sur un brevet avec lui-même dans la base de données SIRU? Nous commençons par examiner l'information disponible dans chaque ensemble de données. Un inventeur a 4 éléments d'information inscrit sur un brevet : son prénom, son nom de famille, sa localisation, i.e. sa ville et son pays, et dans certains cas son état, au moment du dépôt de brevet et le nom du cessionnaire, l'organisation avec laquelle le chercheur collaborait. L'information de financement est toutefois beaucoup plus complète. En effet, elle contient le prénom du scientifique, son nom de famille, son numéro de série, le projet sur lequel il travaillait, l'institution à laquelle il est affilié et son lieu. Ainsi, notre capacité à faire correspondre un chercheur de notre base de données de brevetage avec lui-même dans notre base de données de financement est quelque peu limitée. Il y a deux défis principaux à relever lorsque vient le temps de faire correspondre les scientifiques à eux-mêmes : (1) l'homonymie et (2) la synonymie. Calero *et al.* (2006) étaient confrontés à un problème similaire lorsqu'ils tentaient d'identifier certains groupes de recherche en effectuant l'analyse de publications dans le domaine de la nanotechnologie. Afin de résoudre le problème d'homonymie, c'est-à-dire la probabilité que 2 ou plusieurs chercheurs possèdent le même nom, ils créent une combinaison auteur/organisation. Ceci ne nous serait cependant pas utile par contre car nos données couvrent près de 3 décennies et la plupart des chercheurs ont ainsi eu beaucoup de temps pour travailler à différentes firmes et universités et cette méthode introduirait un supplément de bruit dans notre analyse. Afin de limiter l'impact du problème de synonymie, Calero *et al.* (2006) limitent leurs données à seul pays et gardent les combinaisons avec 6 publications et plus. Notre problème de synonymie est limité par 2 facteurs. (1) nos données sont restreintes au Canada pour les brevets et au Québec pour le financement. (2) dans la préparation des données nous combinons l'ensemble des informations provenant de SCOPUS, les trois conseils subventionnaires canadiens (OCTROI)<sup>16</sup>,

---

<sup>16</sup> La base de données du Conseil Subventionnaire Canadien construite par l'Observatoire des Sciences et des Technologies (OST) contient de l'information sur le financement gouvernemental de chercheurs canadiens par le Conseil de Recherche en Sciences Naturelles et Génie du Canada (CRSNG), les Instituts de Recherche en Santé du Canada (IRSC) et du Conseil de Recherche en Sciences Humaines du Canada (CRSH). Même si ces bases de

avec celles de l'USPTO. En fait, tous les noms de chercheurs ont été examinées un à un et ont été assignées un numéro d'identification selon une procédure spécifique<sup>17</sup>. Ceci nous assure que, peu importe l'orthographe, un chercheur qui épelle son nom de manière différente lors de la publication d'un article, l'application d'un brevet ou d'une demande de subvention serait identifié comme la même personne. Lorsque nous appliquons cette stratégie à la base de données SIRU, nous couvrons la majorité des noms de scientifiques, ainsi réduisant nos chances d'omettre la correspondance de certains noms<sup>18</sup>.

### 3.4.3 Identification et assignation des cessionnaires

Chaque brevet possède un ou plusieurs cessionnaires qui détiennent les droits d'exploitation sur leur invention. Dans notre cas, un des grands défis est de faire correspondre les noms des organisations pourvoyeur de fonds avec les noms des cessionnaires de brevet de nanotechnologie et de biotechnologie. À première vue, cette procédure semble triviale, cependant ce n'est pas le cas. Même si le problème d'homonymie n'est pas aussi important que pour l'association des chercheurs, le problème de synonymie est très présent. Afin de nous assurer que tous les cessionnaires aient le numéro d'identification approprié, nous parcourons chacun des 49172 données de la table de cessionnaires-brevets et attribuons manuellement le numéro d'identification approprié aux cessionnaires. Cet exercice tient compte des restructurations d'entreprises, variations de noms (incluant français à anglais) et des fusions horizontales connues, mais il est important de mentionner que la plupart de l'histoire des entreprises liées à des PME

---

données ne sont pas directement nécessaires pour les besoins de notre analyse, leur utilisation augmente la précision de notre procédure de correspondance des chercheurs et ils seront utiles dans des analyses ultérieures.

<sup>17</sup> Les chercheurs ayant le même nom de famille et le même prénom ont été identifiés comme la même personne. Nous avons fait de même avec les initiales. Cependant lorsque des initiales laissent entrevoir de l'ambiguïté, nous nous sommes fiés aux affiliations du chercheur pour l'identifier correctement.

<sup>18</sup> Une évaluation du taux de faux-positifs et de faux-négatifs de cette méthode serait recommandable. Mais, étant donné le nombre de chercheurs étudiés, cette tâche nécessiterait une analyse qualitative profonde pour avoir un taux représentatif de notre base de données.

privées est quelque peu limitée. Ayant proprement identifié chaque cessionnaire, nous faisons correspondre les noms des cessionnaires avec les organisations pourvoyeurs de fonds dans la base de données SIRU en utilisant une recherche de motif simple. Cette méthode donne 7465 instances de collaboration. Afin d'améliorer nos résultats, nous ajoutons 226 noms alternatifs (`link_assignees_siru`). Nous ignorons aussi les mots vides et autres textes superflus, i.e. « corporation », « incorporé », « limited », etc. en nous assurant que des noms communs ne seraient pas faussement associés. Par exemple, « Whitaker Corporation » n'a aucune affiliation avec la « Whitaker Foundation », ainsi en enlevant corporation de la recherche nous identifierions faussement la Whitaker Foundation comme une institution qui a tant contribué des fonds de recherche en biotechnologie qu'un cessionnaire de brevet dans le même domaine. En tenant compte de ceci, notre recherche de motif livre 10891 cas de collaboration potentielle avec un total de 360 pourvoyeurs de fonds différents et 1539 chercheurs individuels sur près de 23 ans, c'est-à-dire de 1983 à 2005.

### 3.4.4 Bouclage

Ayant identifié et fait correspondre (1) les inventeurs et les universitaires des bases de données SIRU et USPTO, (2) les cessionnaires et pourvoyeurs de fonds et (3) les inventeurs-universitaires avec les cessionnaires-pourvoyeurs de fonds, nous pouvons facilement identifier les boucles d'innovations. Une fois la tâche de création des boucles effectuée, nous nous concentrons sur la révision du contenu de chacune des boucles afin de nous assurer qu'elles représentent réellement des instances où un financement a mené à un brevetage. Nous examinons le contenu de chaque boucle individuellement en gardant toutes les boucles potentielles, incluant celles qui peuvent avoir un délai de 20 ans entre le financement et le brevetage. Nous faisons ceci pour 2 raisons : (1) seule une analyse qualitative approfondie peut révéler si oui ou non un fonds donné à un chercheur a conduit au brevet, et (2) même s'il est intuitif de croire qu'il ne peut y avoir une courte période de temps entre le financement et l'application du brevet, Daim *et al.* (2007) trouvent que la moyenne se situe entre 5 et 6 ans. Gardant toutes les boucles nous aide aussi à éviter de faussement déclarer une boucle comme étant non-pertinente, en fait, dans la plupart des

cas, il est pratiquement impossible de déterminer si les fonds de recherche et les brevets sont liés en se fiant au contenu descriptif de chacun<sup>19</sup>.

Il est d'ailleurs important de mentionner que même si la majorité de nos données proviennent d'une source publique (i.e. l'USPTO), nous avons à notre disposition de l'information qui est de par sa nature confidentielle : les noms des chercheurs recevant des fonds de recherche et de manière plus importante, les montants attribués à ces chercheurs. Par souci éthique, nous effectuons la censure de leurs noms et des montants reçus.

### 3.4.5 Degré d'application

Nous avons à notre disposition une classification de journaux par *CHI Research* selon leur degré d'application. La classification possède une distinction entre le domaine biomédical et toutes les autres disciplines. Dans le premier cas, les scores correspondent aux définitions suivantes du contenu des journaux :

- 1= observation clinique
- 2= observation clinique et investigation
- 3= investigation clinique
- 4= recherche biomédicale fondamentale.

Dans le deuxième cas, la correspondance est la suivante:

- 1= technologie appliquée
- 2= science de l'ingénierie et science technologique
- 3= recherche appliquée et recherche fondamentale ciblée
- 4= recherche fondamentale

Le degré d'application est un concept que nous empruntons d'Hamilton (2003), mais que nous adaptons à nos propres besoins. En considérant qu'il y a une connaissance spécifique associée à

---

<sup>19</sup> Un exemple classique serait un contrat se nommant « Ribosome Studies » qui ne semble avoir aucun lien avec « Catalytic DNA ».

un brevet particulier, qui dans plusieurs cas sera abordé dans des articles scientifiques par leurs inventeurs (Murray & Stern, 2004), il est possible pour nous de déterminer le degré d'application des connaissances associées au brevet. Premièrement, nous faisons correspondre les inventeurs de brevets de la base de données USPTO avec les auteurs dans la base de données SCOPUS en utilisant une procédure similaire à celle utilisée lors du bouclage. Nous calculons ensuite le degré d'application moyen de chaque journal dans lequel un inventeur a publié durant les 3 années menant à l'application de brevet. C'est ce que nous considérons le degré d'application des connaissances brevetés. Ceci nous donne une indication quant à l'état de la recherche dans le laboratoire pendant la période menant à la demande de brevet. Nous devons mentionner que cette méthode, c'est-à-dire l'utilisation du degré d'application des revues comme mesure de l'application des connaissances incluses dans les brevets, n'a pas été recensée dans la littérature. En effet nous en faisons ici une utilisation nouvelle. Il serait donc prévoyant de l'utiliser avec prudence<sup>20</sup>.

### 3.4.6 Classification des brevets en sous-catégories

Nous pouvons identifier 3 méthodes pour réaliser cette tâche : (1) lire les brevets et assigner chacun à la catégorie à laquelle il devrait appartenir; (2) catégoriser les mots clefs utilisés lors de l'extraction des brevets pour ensuite réaliser une recherche et association simple par motif. (3) utiliser des procédures de forage de texte (*Text Mining*); Voici une description des méthodes de triage possibles.

#### 3.4.6.1 Assignment manuelle

Lire chacun des brevets implique qu'il faut consacrer une période de temps non-négligeable au processus de catégorisation. En effet, même si cette méthode offre à première vue le meilleur

---

<sup>20</sup> Deux facteurs sont à considérer : (1) considérant que le degré d'application est représentatif de l'application des connaissances des brevets, il se peut que des chercheurs publient dans des revues qui sont hors de leur domaine ou dans lequel ils ne publieraient pas habituellement. (2) La mesure des connaissances publiées dans les articles scientifiques pourrait aussi être biaisée par un manque de représentativité des connaissances contenues dans les brevets du même chercheur, e.g. parce que le contenu n'est pas similaire. Mais ceci semble être négligeable, tel que soulevé par Murray et Stern (2004).



taux de précision, minimisant les faux-positifs et les faux-négatifs, classer manuellement 16 419 brevets représente une mauvaise utilisation du temps qui pourrait être utilisé à meilleur escient. En prenant un exemple mentionné par Teichert & Mittermayer (2002), il a fallu une demi-année pour classer une série de 5 188 brevets en 5 catégories technologiques. À ce rythme, nous nous rendons compte que la catégorisation manuelle de nos 16 419 brevets prendrait 1 an et 7 mois! D'ailleurs Liu et al. (2009) mentionnent que se fier à un expert humain pour classer de larges volumes de données est non seulement onéreux en termes de temps et de ressources, mais aussi non fiable si les experts en question sont chargés avec des tâches aussi routinières et minutieuses.

### 3.4.6.2 Recherche de motif simple

La deuxième méthode consiste à classer les termes utilisés pour l'extraction des brevets. Cette méthode manque de complexité et semble intuitivement plus ou moins fiable. En catégorisant les mots clefs selon leur domaine, nous effectuons une recherche de motif en créant une requête SQL qui assigne chaque brevet contenant un mot clef à son domaine respectif. Cette méthode, même si elle a été utilisée à certaines étapes préliminaires, est simpliste. En effet, elle a tendance à classer les brevets dans plus d'une catégorie car elle ne tient pas compte de la fréquence des mots dans un brevet, ni dans l'ensemble des documents.

### 3.4.6.3 Forage de texte

Les deux méthodes ci-dessus n'étant pas robustes, nous choisissons la troisième option : celle du *forage de données*. Le *forage de données* peut être défini comme un ensemble de processus qui permettent d'analyser de grandes quantités d'informations afin d'y trouver des corrélations dissimulées ou des informations nouvelles (Termium, 2010; Weiss et al., 2005). C'est une méthode est utile pour l'analyse dans des domaines de haute technologie ainsi qu'à l'étude des domaines eux-mêmes (Toivonen et al., 2005; Domeniconi et al., 2005; Dong et al., 2005; Shapira & Youtie, 2006; Fall et al., 2003; Li et al., 2007). Dans notre cas, les données sont en format texte, ce qui nous amène à effectuer spécifiquement du *forage de texte*.

Le forage de texte est réalisé à l'aide du logiciel *RapidMiner 5.1.004*. Les raisons pour lesquelles nous avons choisi cet outil sont sa facilité d'apprentissage et sa puissance d'analyse. Une fois le logiciel installé avec la configuration de forage de texte incluse, nous procédons à l'analyse. En prenant la nanotechnologie comme échantillon cobaye nous créons le processus (exemple de

code XML en annexe) qui permet de trier tous les brevets dans leur(s) sous-catégorie(s) respective(s). La procédure suivie est la suivante :

1. Prendre un échantillon aléatoire représentatif (10%) des données et faire l'assignation manuelle de chaque brevet à une catégorie. Le logiciel ne pouvant pas supporter l'importation et l'analyse de toute l'information sur les brevets, nous avons limité l'étude aux 500 premiers caractères des résumés. Cette procédure, similaire à celle utilisée par Fall et al. (2003), permet de normaliser la longueur des documents tout en conservant un haut taux de précision.
2. Procéder au traitement de texte en utilisant le vecteur de mot TF-IDF<sup>21</sup> (*Term Frequency-Inverse Document Frequency*) dont l'équation est la suivante :

$$tf - idf(j) = tf(j) * idf(j)$$

ayant

$$idf(j) = \log\left(\frac{N}{df(j)}\right)$$

Où j : terme quelconque

N : nombre total de documents (brevets)

df(j) : nombre total de documents contenant le terme j

- a. Récupérer les données dans RapidMiner, mettre tout le contenu en minuscules et procéder à une segmentation en mots du contenu entré.
- b. Filtrer les mots vides, e.g. les conjonctions de coordination, de subordination etc., qui alourdissent les processus sans contribuer à la précision de la catégorisation.
- c. Effectuer la lemmatisation des mots avec la méthode de Porter. Ceci permet de réduire la longueur des mots tout en gardant leur signification pertinente. Nous

---

<sup>21</sup> Plusieurs vecteurs peuvent être utilisés (Fréquence de terme, Occurrence de terme, Occurrence de terme binaire), mais la TF-IDF est non seulement la plus répandue (Aizawa, 2003) mais aussi la meilleure de celles disponibles car elle permet de tenir compte de la fréquence intra et extra document.

utilisons la méthode de Porter<sup>22</sup> car elle adaptée spécifiquement pour les mots anglais (n'oublions pas que les informations relatives aux brevets proviennent de la USPTO).

- d. Filtrer les mots en ne gardant que ceux ayant 2 caractères ou plus.
3. Effectuer un « apprentissage » sur la classification insérée. Cette opération permet au logiciel d'apprendre comment classer les documents qui seront prochainement introduits dans le programme. Dans un premier temps, nous employons la méthode de classification des k plus proches voisins (K-NN)<sup>23</sup> utilisant la similarité cosinus comme mesure principale. La similarité cosinus représente une mesure de similitude entre deux documents et est souvent utilisée en fouille de données, plus particulièrement en comparaison de documents en fouille de texte (Grossman, 1998). Ayant la requête Q, pour le document D<sub>i</sub> et les poids w établies, dans ce cas-ci par le TF-IDF, son équation est la suivante:

$$SC(Q, D_i) = \frac{\sum_{j=1}^t w_{qj} d_{ij}}{\sqrt{\sum_{j=1}^t (d_{ij})^2 \sum_{j=1}^t (w_{qj})^2}}$$

4. Effectuer une validation de l'apprentissage précédente. Il faut valider la méthode d'apprentissage pour vérifier qu'elle rende une classification appropriée et pour ce faire, nous employons la fonction de performance intégrée à la fonction de validation croisée

---

<sup>22</sup> L'algorithme de la méthode de Porter se trouve directement dans le site de l'auteur à l'adresse suivante : <http://tartarus.org/~martin/PorterStemmer/java.txt>

<sup>23</sup> Il existe plusieurs méthodes de classification/régression. Entre autres, nous pouvons utiliser les méthodes de Lazy Modeling, Modélisation Bayésienne, d'arbres décisionnels, réseaux de neurones, induction de règle, etc. De nombreuses études (Teichert & Mittermayer, 2002; Li et al., 2007; Fall et al., 2003) effectuent l'évaluation de la performance des méthodes, et il semble clair qu'aucune méthode spécifique est universellement meilleure que les autres, tel que soulevé par le théorème du « No Free Lunch » (Wolpert & Macready, 1997). Le choix de la méthode de classification dépend donc du cas.

(X-Validation). Pour trouver un compromis entre qualité de la validation et vitesse d'opération nous fixons la validation à 5 itérations avec échantillonnage aléatoire.

5. Récupérer et traiter les données à classer. Nous répétons ici l'étape 2 a, b et c avec les données non-triées.
6. Appliquer le modèle aux nouvelles données.

Il y a une grande variété dans la gamme des méthodes d'apprentissage. Certaines sont plus robustes que d'autres, mais la robustesse n'est pas toujours synonyme de précision. C'est pour cela que nous procédons avec deux autres méthodes d'apprentissages, soient Naive Bayes et Support Vector Machine (SVM) pour déterminer laquelle est mieux adaptée aux besoins de notre analyse.<sup>24</sup>

Pour ce faire nous prenons un échantillon aléatoire de données avec lequel nous répétons la même analyse mais à chaque fois utilisant une technique d'apprentissage différente. Nous révisons attentivement le contenu de 350 brevets et attribuons à chacun la catégorie appropriée. Nous comparons ensuite les taux de précision dans la classification entre elles pour garder la meilleure.

D'ailleurs, plusieurs méthodes de combinaison recensées par Kotsiantis (2007) permettent d'améliorer la précision de la classification. Elles incluent l'empilage, la généralisation en cascade, arbres méta-décisionnelles (MDT), arbre Naive Bayes flexible, etc.<sup>25</sup> Nous décidons d'employer la méthode par vote qui consiste à compiler les résultats de chaque méthode et

---

<sup>24</sup> Nous devons aussi mentionner que nous avons exécuté plusieurs instances d'optimisation afin d'avoir des paramètres qui produisent le meilleur taux de précision. Ces optimisations de paramètres se trouvent en annexe et couvrent les éléments principaux tels la valeur de K pour K-NN, l'application ou non de la correction de Laplace pour Naive Bayes et le type de noyau pour SVM.

<sup>25</sup> La méthode utilisée par Teichert & Mittermayer(2002), i.e. considérer un brevet catégorisé que lorsque toutes les méthodes indiquent le même résultat donne un rendement très intéressant. Cependant, ce type de combinaison exclurait plusieurs des résultats de notre analyse, nous optons donc de ne pas l'utiliser.

d'assigner la prédiction à la catégorie majoritaire (Kotsiantis, 2007), en n'utilisant que les méthodes qui ne sont pas systématiquement surclassées par les autres. En utilisant les paramètres optimisés comme base de modélisation, les résultats de précision sont les suivants:

Tableau 3-1: Précision de la classification selon la méthode d'apprentissage

		K-NN	Naive Bayes	SVM	Vote
Nanotechnologie					
	Métrologie & processus	86,79%	84,47%	86,72%	87,01%
	Électronique et dispositifs	77,09%	70,07%	74,33%	76,02%
	Structures et matériaux	83,30%	80,83%	81,93%	84,42%
Biotechnologie					
	Tout	86,00%	84,50%	78,50%	93,00%

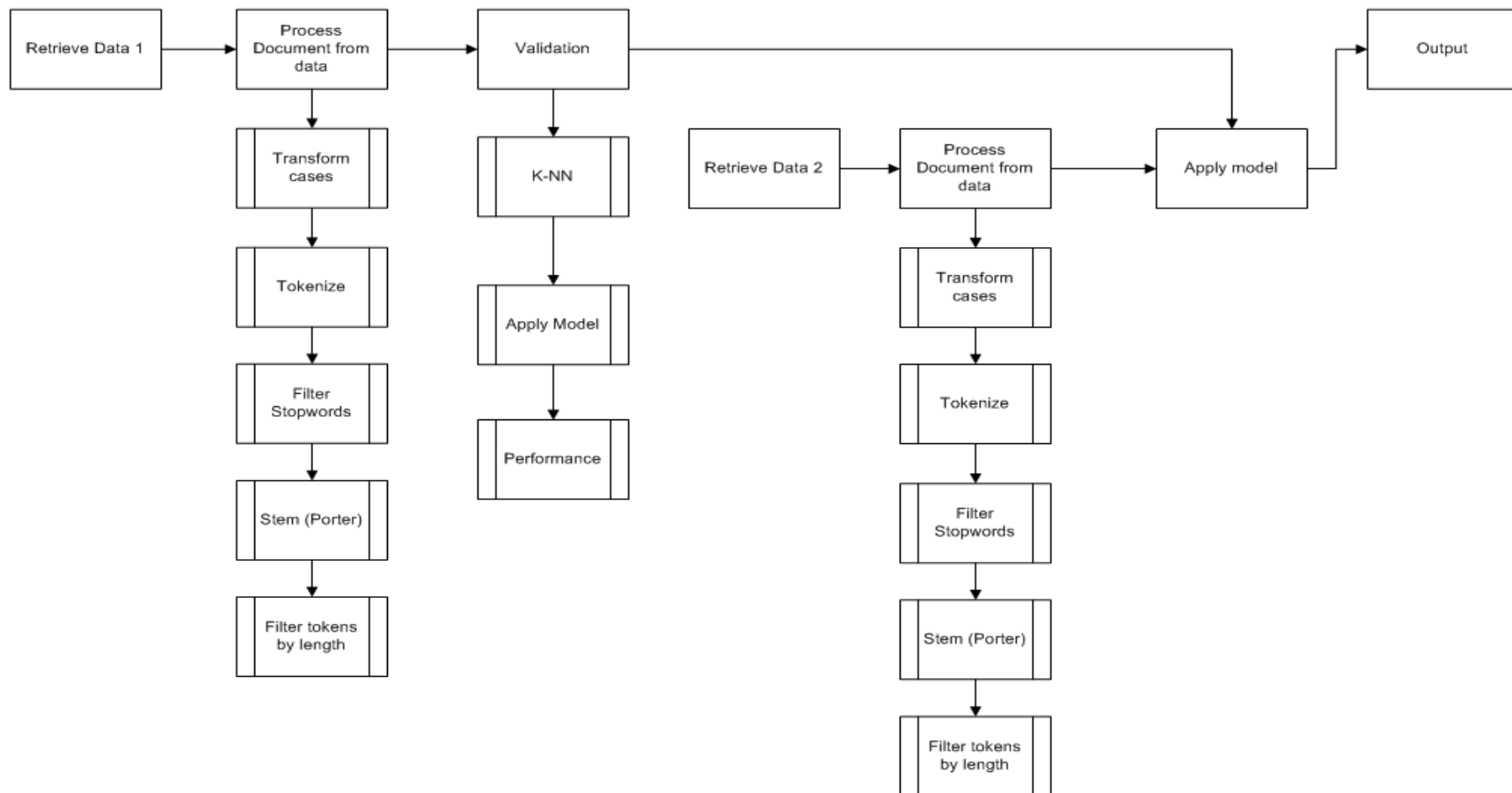


Figure 3-2: Méthode de catégorisation des brevets dans Rapidminer

La Figure 3-2 illustre la séquence des opérations employée dans la catégorisation des brevets et décrite plus haut. Les sous-processus contenus dans les opérateurs imbriqués sont listés en dessous de ceux-ci (dans les rectangles intégrés). L'optimisation des paramètres

et la combinaison des classifications se fait dans un fichier différent pour faciliter la faisabilité procédurale; la séquence des opérations qui décrit cette méthode se trouve dans la Figure 3-3.

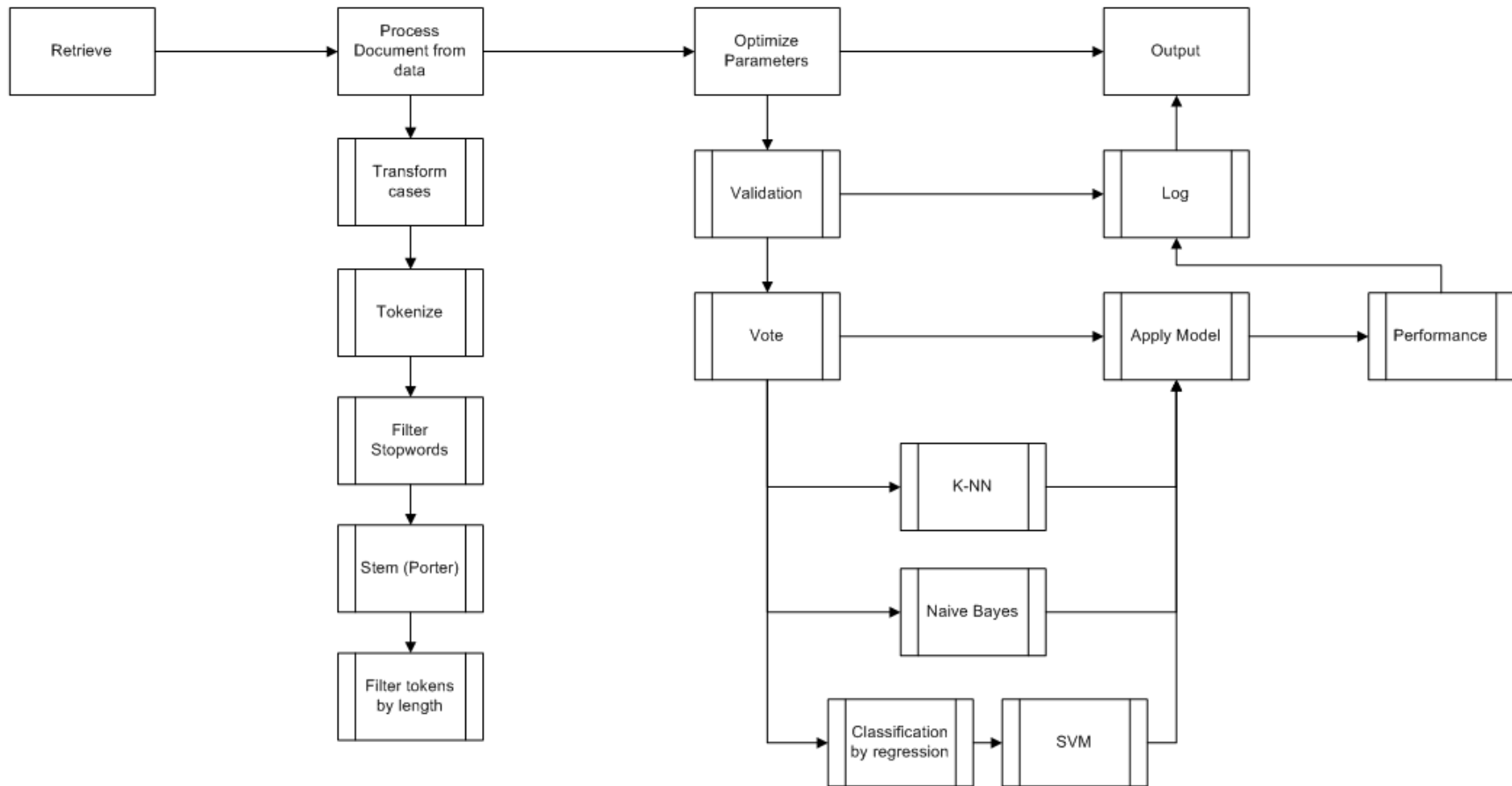


Figure 3-3: Méthode employée dans l'optimisation des paramètres et la combinaison des classifications dans Rapidminer

Selon le Tableau 3-1, La méthode de Naive Bayes semble être la moins bien adaptée pour prédire la catégorie des brevets en nanotechnologie et celle de SVM est la moins précise pour les brevets en biotechnologie. Il semble que la combinaison par vote soit la meilleure option. Voici, la classification des brevets par catégorie utilisant cette méthode :

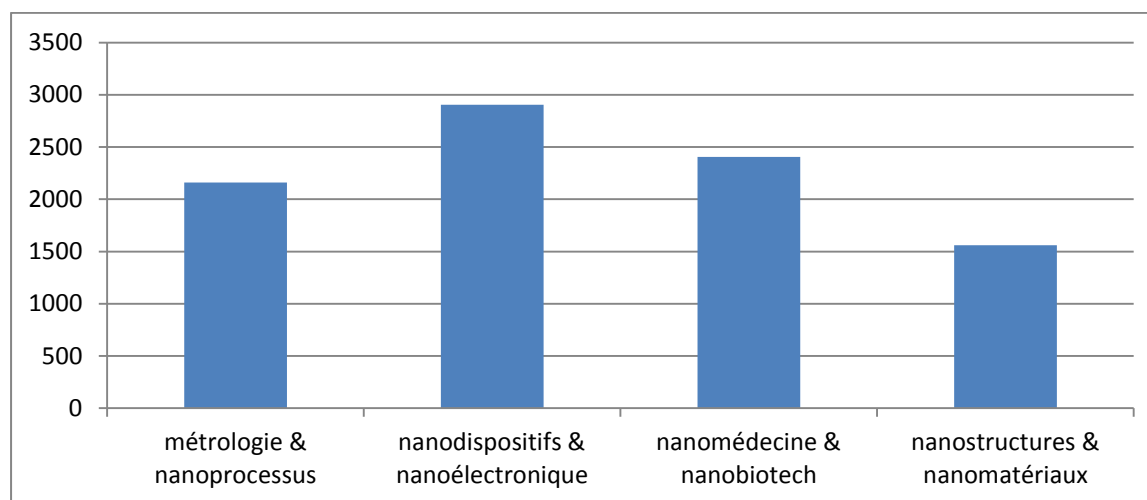


Figure 3-4: Classification de brevets nanotechnologiques<sup>26</sup>

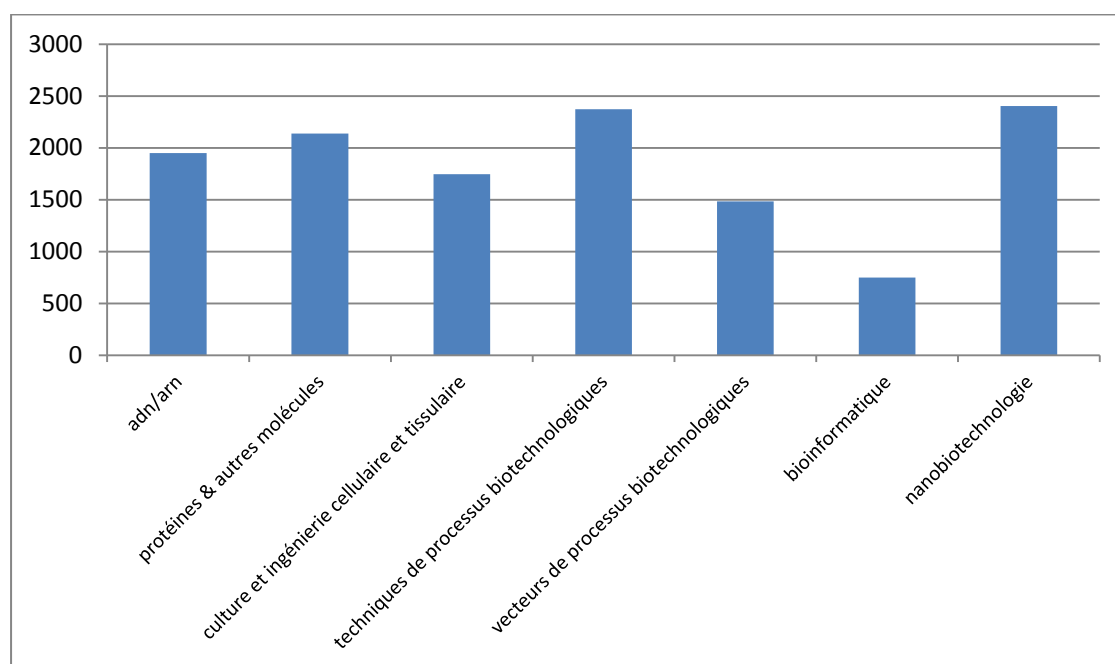


Figure 3-5: Classification de brevets biotechnologiques

<sup>26</sup> Les brevets en nanotechnologie peuvent se trouver dans plus d'une catégorie à la fois.



Ceci est bien, toutefois nous réalisons qu'une des grandes faiblesses de cette méthode réside dans le fait que son efficacité dépend de la présence de résumés de brevets clefs dans l'échantillon initial utilisé lors de l'étape 1. En effet, plus l'échantillon initial est représentatif de l'ensemble des données, mieux sera la classification résultante. La plupart des études revues, utilisent un ratio de 1 pour 2 dans la classification, c'est-à-dire 2 brevets à classer pour chaque brevet utilisé dans l'apprentissage. Pour nous, ce ratio est beaucoup plus petit en nanotechnologie (10%) et infiniment plus petit en biotech (N/A). Par contre le coût associé à une augmentation du nombre de brevets utilisés lors de l'apprentissage semble dépasser l'amélioration potentielle de la classification qui en résulterait.

### 3.4.7 Méthodologie Statistique Comparative

Dans la première partie du projet, nous menons des comparaisons statistiques qui permettent de valider les hypothèses 1 à 4, i.e. d'identifier des différences significatives entre les domaines de la nanotechnologie, de la biotechnologie. Étant donnée la gamme des indicateurs étudiés et la diversité du type de données (e.g. dichotomiques, continues, etc.) nous ne pouvons nous limiter à un seul test, car les hypothèses ou conditions d'applications ne se trouvent pas toujours à être remplies. Voici un survol des tests employés et de la démarche empruntée pour les appliquer.

#### ANOVA

Le test de choix est celui de l'ANOVA ou *ANalysis Of VAriance*. Il permet de déterminer s'il y a une différence significative entre les groupes à l'étude, i.e. il révèle si les groupes proviennent de la même population. Le test consiste à valider l'hypothèse  $H_0$  contre l'hypothèse  $H_1$  :

$H_0$  : Les moyennes des différents groupes sont égales.

$H_1$  : Au moins une moyenne des groupes à l'étude n'est pas égale aux autres.

Deux conditions doivent exister avant de pouvoir employer cette méthode. Premièrement, il doit y avoir normalité des distributions. Cette normalité est vérifiée à l'aide du test de Skewness-Kurtosis. Ceci est fait par l'intermédiaire de la procédure « `sktest {nom de variable}, noadjust` » dans Stata. L'option « `,noadjust` » permet de supprimer l'ajustement de Royston (1991) effectué

sur le  $\chi^2$  global et sa significativité (Stata, 2009). Ainsi si les données ne présentent pas de tendance normale, il faut soit considérer un ajustement pour ramener les skewness et kurtosis près de 0 et 3 respectivement en prenant le logarithme naturel ou la racine carrée de la variable par exemple, ou mettre de côté le test ANOVA en le remplaçant par un test non-paramétrique.

La deuxième condition est celle d'homoscédasticité, c'est-à-dire d'égalité des variances  $s_k$ . Cette égalité peut être vérifiée de plusieurs manières, selon la normalité des données. S'il a été démontré qu'il y a normalité des données avec le test de Skewness-Kurtosis alors nous pouvons procéder avec le test de Bartlett<sup>27</sup>. Ce dernier propose l'hypothèse nulle qu'il y a égalité des variances et l'hypothèse alternative qu'elles sont différentes. Avec des variances  $s_1^2, s_1^2, \dots, s_k^2$  et des degrés de liberté  $v_1, v_2, \dots, v_k$  et  $v_1 = n_i - 1$ , le test de Bartlett possède la statistique suivante (Green & Margerison, 1978):

$$T = \frac{v \ln s^2 - \sum v_i \ln s_i^2}{1 + \left\{ \left( \sum \frac{1}{v_i} \right) - \frac{1}{v} \right\} / \{3(k-1)\}}$$

Où

$$v = \sum v_i$$

$$s^2 = \sum v_i s_i^2 / v$$

Un des avantages de ce test est qu'il est généré automatiquement à l'exécution de la procédure ANOVA dans Stata. Si nous nous trouvons dans une situation où le test ANOVA n'est pas de mise, soit par une non-normalité extrême<sup>28</sup>, soit par une hétérogénéité des variances, nous procédons avec des tests non-paramétriques pour déterminer l'égalité des moyennes.

---

<sup>27</sup> Le test de Bartlett's requiert cette normalité qui est en fait un aspect primordial de sa validité (Green & Margerison, 1978)

<sup>28</sup> Le test d'analyse de la variance reste robuste même lorsqu'il y a un écart léger de la normalité.

### Kruskal-Wallis

Le premier test non-paramétrique est celui de Kruskal-Wallis. Il est une généralisation du test de Wilcoxon-Mann-Whitney et permet de comparer K échantillons supérieurs ou égaux à 2 entre eux pour déterminer s'ils sont de la même population (Haccoun & Cousineau, 2010). L'hypothèse nulle  $H_0$  à tester contre son alternative  $H_1$  est la suivante :

$H_0$  : Tous les groupes proviennent de la même population.

$H_1$  : Au moins un groupe ne provient pas de la même population.

Règle de décision : Rejeter  $H_0$  si  $g > g_{\text{critique}}$

Avec  $SR_i$  étant la somme des rangs du groupe  $i$ ,  $N_i$  la taille du groupe  $i$ ,  $N$  la taille de tous les groupes confondus et  $k$  le nombre de groupes, l'équation du test de Kruskal-wallis est la suivante (Haccoun & Cousineau, 2010):

$$g = \frac{\sum_{i=1}^k \frac{SR_i^2}{N_i}}{N(N+1)/12} - 3(N+1)$$

Il est important de mentionner que certains (Ruxton & Beauchamp, 2008) proposent que le test de Kruskal-Wallis ne soit pas un équivalent non-paramétrique du test ANOVA car il testerait l'homogénéité stochastique, i.e. la distribution de la population, plutôt que de comparer une mesure univariée. Cependant, à toutes fins pratiques, nous l'utilisons en vérifiant la similarité des distributions.

### Mann-Whitney (Wilcoxon-Mann-Whitney ou rank-sum)

Le deuxième test non-paramétrique est celui de Mann-Whitney. Il est applicable si et seulement si le nombre de groupes à comparer est de deux et permet de décider si les échantillons de ces deux groupes sont issus de la même population. Il est approprié lorsque la normalité des données est douteuse et lorsqu'il est possible d'ordonner les données de chaque groupe dans un rang précis (Haccoun & Cousineau, 2010). L'hypothèse nulle  $H_0$  à tester contre son alternative  $H_1$  est donc :

$H_0$  : La somme des rangs d'un groupe est égale à la somme des rangs de l'autre groupe.

H1 : La somme des rangs d'un groupe est supérieure à celle de l'autre.

Le test de Mann Whitney s'écrit ainsi (Haccoun & Cousineau, 2010) :

$$z = \frac{|SR_1 - N_1(N + 1)/2| - 0,5}{\sqrt{N_1N_2(N + 1)/12}}$$

Où  $SR_1$  la somme des rangs du groupe 1,  $N_i$  la taille du groupe i, N la taille totale des groupes.

En tenant compte des informations ci-dessus, nous avons créé un arbre décisionnel permettant de choisir la méthode de comparaison appropriée.

Cette procédure (Figure 3-6) a été appliquée pour répondre au premier objectif de recherche, celle de l'analyse des similarités entre la nanotechnologie et la biotechnologie. Elle regroupe la comparaison du nombre de revendications associées aux brevets, celle du nombre de citations et du degré d'application. Les comparaisons subséquentes portent sur le nombre de revendications des brevets et des citations en boucles d'innovation versus celles à l'extérieur des boucles. L'effet de l'intervalle de temps entre le financement et le brevetage est ensuite analysé pour déterminer si cela a un effet sur le degré d'application. Nous n'incluons pas tous les sous-domaines de la nanotechnologie et de la biotechnologie à ce stade de la recherche, car elles sont trop nombreuses et la comparaison une à une serait non-pertinente. Il est à noter que les sous-domaines complets sont intégrés aux modèles économétriques plus loin lorsque le domaine comme tel se trouve à avoir un effet significatif sur la variable dépendante. La liste complète des comparaisons menées et le détail des comparaisons comme tel se trouvent à l'annexe 6.

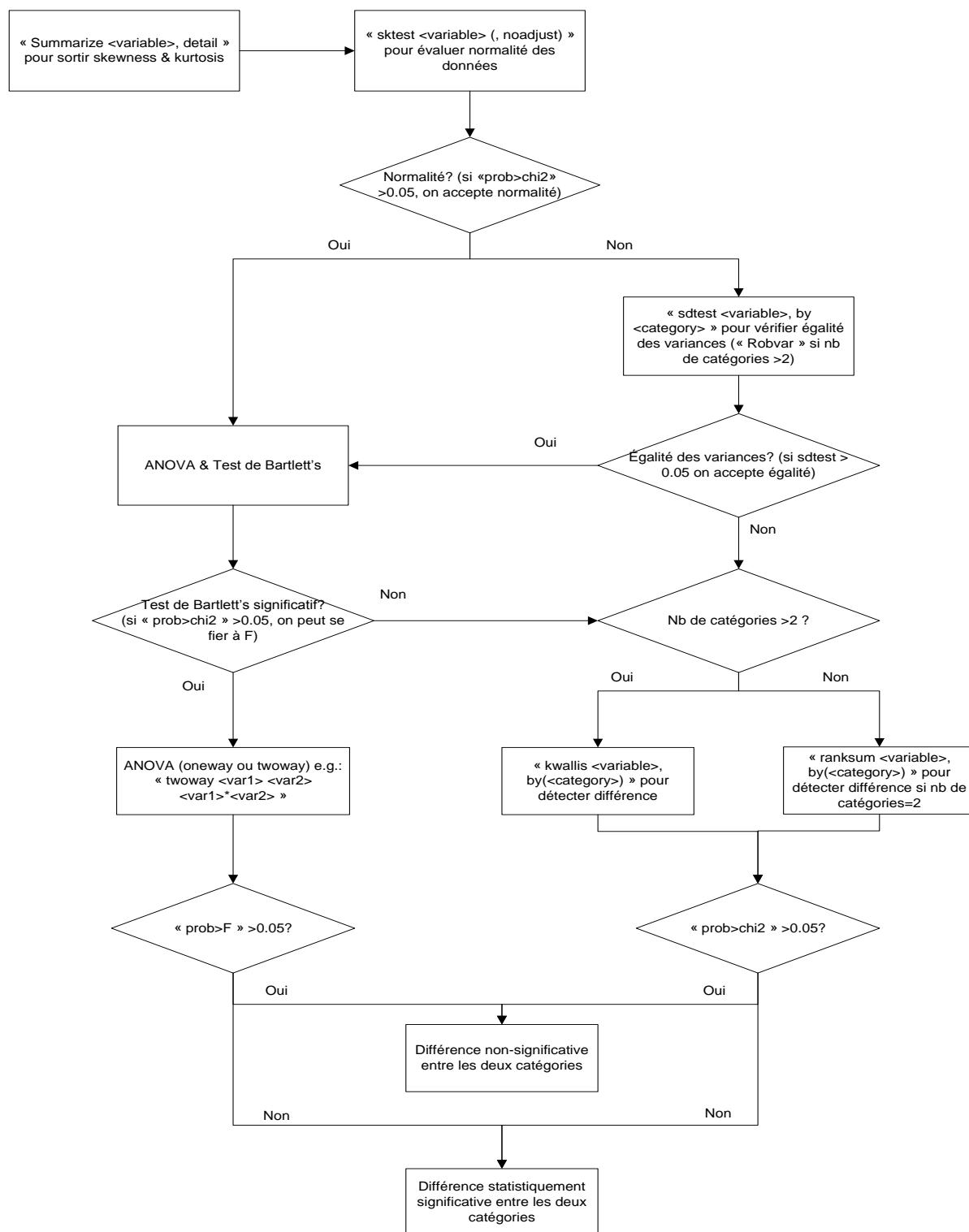


Figure 3-6: Arbre décisionnel pour tester les différences de moyennes et variances

### 3.4.8 Méthodologie Économétrique

#### 3.4.8.1 Variables

Dans cette section nous spécifions les variables employées dans les modèles économétriques. Nous commençons avec les variables que nous tentons d'expliquer, c'est-à-dire les variables dépendantes. Le nombre de brevets sur lequel un chercheur est nommé comme inventeur au temps  $t$  est exprimé par [nbpat]. Pour ce qui est des indicateurs de qualité, le nombre de citations qu'un chercheur reçoit sur une période de 5 ans suivant l'octroi du brevet est exprimé par [nbcitp5]. Le nombre de revendications incluses dans les brevets sur lesquels il est mentionné à l'année  $t$  est représenté dans [nbclaims].

Ces variables sont dépendantes de plusieurs variables exogènes communes, telles celles portant sur le financement du chercheur, ses publications scientifiques, les effets de réseaux qui y sont rattachés, et même les caractéristiques du chercheur comme tel.

Le financement du chercheur est décomposé en deux catégories, les subventions qu'il reçoit [totsubv] et les contrats qu'il reçoit [totcont] au temps  $t$ . Nous adaptons ces deux variables pour représenter plus fidèlement l'importance du financement en calculant le logarithme naturel des moyennes sur des périodes de 3 ans représenté par [ltotsubvmoyan3] et [ltotcontmoyan3]. Afin d'avoir une idée du financement cumulatif d'un chercheur, nous créons [totsubv\_past] et [totcont\_past] qui représentent les totaux des subventions et des contrats qu'un chercheur a reçu durant sa carrière académique. Pour évaluer la magnitude de la recherche qui se fait dans les universités d'affiliation nous considérons aussi les subventions et les contrats du corps professoral de ces universités sur des périodes de 3 ans avec [ltotsubvmoyan3wuu] et [ltotcontmoyan3wuu].

L'analyse des publications d'un chercheur est importante tant dans l'évaluation de sa performance de brevetage que dans l'évaluation de la qualité des brevets. Nous utilisons la variable [nbarticlemoyan3] pour représenter le nombre d'articles que le chercheur a publié sur une période de 3 ans. Tout comme pour le financement, nous compilons aussi des variables pour exprimer la quantité d'articles publiés durant sa carrière [nbart\_past], les moyennes d'articles publiés sur une période de 3 ans pour les chercheurs en nanotechnologie ou biotechnologie à son université d'attache [nbarticlemoyan3wuu]. Une variable importante qui est aussi compilée est

bien sûr le degré d'application des connaissances associées aux brevets sur la même période [deg\_appmoyan3]<sup>29</sup>.

Les effets réseaux étant reconnus comme ayant des effets sur les extrants scientifiques (Zucker & Darby, 1996; Beaudry & Schiffauerova, 2011) nous incluons les variables de cliquishness [lmcliquind23] et de betweenness centrality [ldmbetcent3] sur une période de 3 ans dans les régressions. Elles permettront d'analyser plus en détail l'impact de la position académique d'un chercheur tant sur la quantité que sur la qualité des brevets.

Pour ce qui est des caractéristiques personnelles du chercheur nous employons son [age] qui représente le temps écoulé depuis la première publication de ce-dernier. Il sert de mesure d'expérience académique mais aussi comme mesure pour justifier des variables telles [ltotsubvmoyan3] et [nbarticlemoyan3]. Une autre caractéristique du chercheur est la variable [codechair33] qui est un indicateur du type de financement qu'il a reçu. C'est une variable ordonnée, notre seule en fait, qui vaut 0 si un chercheur n'est pas associé à une chaire de recherche, 1 s'il est titulaire d'une chaire industrielle, 2 pour une chaire du CRSNG ou des IRSC et 3 pour une chaire de recherche du Canada.

Pour ce qui est des sous-domaines de recherche, [nano\_excl] exprime l'instance pour laquelle un chercheur œuvre (publie) simplement en nanotechnologie, i.e. excluant la nanobiotechnologie, idem pour la biotechnologie avec [bio\_excl], [nanobio] représente le travail fait en nanobiotechnologie. Pour ce qui est des brevets comme tels, [excl\_nano\_pat] vaut 1 lorsque le ou les brevets à l'année en cours ne portent que sur la nanotechnologie, [excl\_bio\_pat] vaut 1 lorsque le ou les brevets à l'année en cours ne portent que sur la nanotechnologie et [nanobio\_pat], étant le groupe de référence (donc généralement la variable omise), contient tous les autres brevets. Les sous-domaines des brevets en nanotechnologie sont [metrology\_processus\_pat],[devices\_electronics\_pat],[structures\_materials\_pat] et celles de la biotechnologie sont [dna\_rna\_pat], [proteins\_molecules\_pat], [cell\_tissue\_engineering\_pat], [process\_biotech\_pat], [gene\_rna\_pat], [bioinformatics\_pat]. Finalement, la catégorie commune aux deux sous-ensembles est [nanobio\_pat] qui représente les brevets exclusivement contenus en nanobiotechnologie.

---

<sup>29</sup> Elle est aussi compilée de manière ponctuelle avec [deg\_app0]

Deux autres variables modélisées dans le cadre de ce travail sont les universités d'affiliation du chercheur [univ] et les années [years]. Elles permettent entre autres de vérifier l'impact d'appartenance à une institution universitaire particulière, e.g. Université Laval [d01Laval], Université McGill [d02McGill], Université de Montréal[d04MtlG], Université Concordia [d07Concordia], Université de Sherbrooke regroupé avec l'Université Bishop [d08SherbrookeG], et l'Université du Québec regroupée [d09UQAMG], ou d'un effet temporel avec [d1996-d2005].

Il est à noter que d'autres variables secondaires, telles les variables d'interaction<sup>30</sup>, sont compilées, mais considérant leur importance moindre pour l'instant, elles ne seront pas explicitées ici. La liste complète des variables se trouve à l'annexe 23.

### 3.4.8.2 Modèle de Poisson et à Binomiale négative

Le Modèle Linéaire Classique (MLC) utilisant la méthode des moindres carrés ordinaires (MCO) est le modèle économétrique de base et nous l'utilisons lorsque ses conditions<sup>31</sup> sont respectées. Cependant, nous avons affaire à des données de comptage (discrètes) comme variables dépendantes. Ainsi, pour ces cas la méthode des MCO n'est pas adéquate car certaines de ses hypothèses de base telles l'homoscédasticité<sup>32</sup> et la normalité de la variable dépendante ne sont pas respectées (Wooldridge, 2000). Les modèles que nous estimons sont donc plutôt des régressions de Poisson simple et à Binomiale Négative.

Le modèle de Poisson est la vedette des régressions de données de comptage. Ayant  $y_i$ , une variable discrète non-négative, qui est une variable aléatoire indépendante et suivant une loi de Poisson de paramètre  $\lambda_i$  nous avons :

---

<sup>30</sup> Une variable d'interaction (variable interactive) est utilisée pour considérer l'effet combinatoire de deux ou plusieurs variables. Lorsque nous soupçonnons une interaction entre deux variables, il est pertinent de modéliser l'interaction.

<sup>31</sup> Le modèle des moindres carrés ordinaire doit respecter les hypothèses suivantes lorsque  $y = a + bx + \mu$ . H1 :  $y$  et  $x$  sont observables avec  $y$  et  $\mu$  étant aléatoires. H2 :  $E(\mu) = 0$ ,  $i = 1, \dots, n$ . H3 :  $\mu$  sont indépendants et identiquement distribués (iid). H4 :  $(X'X)$  est une matrice singulière. H5 :  $\mu \sim n(\sigma, \sigma^2 I_n)$ .

<sup>32</sup> Distribution identique des erreurs  $\mu$ .



$$p(y_i = k) = \frac{e^{-\lambda_i} \lambda_i^k}{k!}$$

À la condition que  $E(y_i) = \text{Var}(y_i) = \lambda_i$

Wooldridge (2000) mentionne que lorsque nous traitons des données de comptage, il y a peu de raisons de vouloir utiliser un modèle autre que celui de Poisson. Un cas d'exception cependant serait lorsque la variance n'est pas égale à la moyenne. Dans ce cas le modèle à binomiale négative est préférable. La binomiale négative<sup>33</sup> est une version non-restrictive, i.e. généralisée, du modèle de Poisson qui inclut un effet individuel non-observé dans la moyenne conditionnelle (Greene, 2002).

Afin de choisir entre la régression à binomiale négative et la régression de Poisson, nous devons évaluer s'il y a équidispersion ( $\text{Var}(y_i) = E(y_i)$ ) ou surdispersion ( $\text{Var}(y_i) > E(y_i)$ ). Cameron & Trivedi (1986) ont compilé plusieurs méthodes pour tester la surdispersion. Ayant,  $E(e^{u_i}) = 1$ ,  $\text{Var}(e^{u_i}) = \eta^2$ , il faut évaluer la probabilité d'avoir  $\eta^2$  différent de 0 et positif.

Lorsqu'il y a surdispersion nous devons employer la régression à binomiale négative (Gagnon, 2010). En termes pratiques, la valeur de  $\eta^2$  est compilée en générant une régression binomiale négative dans Stata et est représentée par la valeur *alpha*. C'est en examinant son degré de signification<sup>34</sup> que nous pouvons déterminer si elle est significative ou non.

La base de données employée pour cette analyse a été montée en panel. Le panel est utilisé lorsque la modélisation de deux dimensions (dont une est le temps) est nécessaire. Ceci permet de maintenir un lien entre les individus, i.e. un lien horizontal, mais aussi un lien pour le même individu dans le temps, i.e. un lien vertical, et donne donc une qualité supplémentaire à la modélisation. Le dilemme en prenant un modèle en panel vient alors du choix d'effets

---

<sup>33</sup> Sa fonction de densité se trouve ci-dessous. La fonction gamma est utilisée pour fins de commodité mathématique (Greene, 2002).

$$f(y_i|x_i) = \int_0^\infty \frac{e^{-\lambda_i u_i} (\lambda_i u_i)^{y_i}}{y_i!} \frac{\theta^\theta u_i^{\theta-1} e^{-\theta u_i}}{\Gamma(\theta)} du_i$$

<sup>34</sup> Le degré de signification représente la probabilité d'obtenir un résultat déterminé en supposant que l'hypothèse nulle est vraie.

spécifiques. Il est possible de modéliser des données de panel avec soit (1) des effets fixes ou (2) des effets aléatoires.

- (1) Le modèle à effets fixes suppose que les termes d'erreurs associés aux individus et au temps sont des valeurs non-aléatoires. Un problème relié au modèle à effets fixes est l'*Incidental Parameter Problem* (IPP) qui fait croître le nombre de degrés de liberté par 1 à chaque fois qu'une observation s'ajoute.
- (2) Le modèle à effets aléatoires ne supporte pas la résolution par Moindres Carrés Ordinaires. Cependant, il se débarrasse du IPP et nous permet à la fois de considérer de l'hétérogénéité modélisée et non-modélisée. Il est ainsi préférable aux effets fixes, car il permet de faire une inférence applicable à la population et non seulement à l'échantillon (Gagné, 2010).

Afin de décider quels effets employer, nous nous fions à deux tests (1) le test de Breusch-Pagan et (2) le test de spécification d'Hausman.

- (1) Le test de Breusch-Pagan détecte les effets spécifiques aléatoires. Sous l'hypothèse nulle, nous rejetons la présence d'effets aléatoires.
- (2) Le test de spécification d'Hausman évalue s'il y a un lien entre les variables indépendantes et les effets aléatoires. Sous l'hypothèse d'orthogonalité, les deux estimateurs sont équivalents (Gagnon, 2010). Un exemple de démarche de ces deux tests avec Stata se trouve à l'annexe 22.

Lorsque cela est possible nous spécifions le panel dans nos régressions soit en Poisson, soit en binomiale négative. Lorsque cela n'est pas possible, par exemple lorsqu'il y a trop peu d'observations par groupe, nous utilisons la sous-procédure un « vce (cluster id) ». Celle-ci précise que les erreurs standards permettent la corrélation intra-groupe, assouplissant ainsi l'exigence habituelle que les observations soient indépendantes (Stata, 2010). Cela est évidemment moins robuste que d'effectuer les régressions directement en panel, mais reste une option considérablement supérieure à simplement négliger les interactions entre individus.

### **3.4.8.3 Modèle à binomiale négative à zéro-augmenté**

Dans certains cas de modélisation de données de comptage, il se peut qu'il ait un excès de zéros. Cet excès de zéros est souvent dû au fait que les zéros sont engendrés par plus d'un facteur. De

manière un peu plus concrète, prenons le nombre de brevets que possèdent les chercheurs universitaires comme exemple. Un chercheur à une année  $t$  pourrait ne pas avoir de brevet parce qu'il n'avait pas à sa disposition l'infrastructure nécessaire pour développer une invention brevetable, ou parce que même en ayant tous les outils et le personnel requis, il n'arrivait simplement pas à breveter une invention pour une raison quelconque.

Pour évaluer s'il y a un excès de zéros et donc décider si le modèle à binomiale négative à zéro-augmenté est plus adéquat que la binomiale négative ordinaire, nous procédons au test de Vuong. Un test  $z$  compare un modèle à l'autre. Si la valeur de  $z$  est significative, cela indique que le modèle à binomiale négative à zéro-augmenté est préférable.

À la Figure 3-7 Ci-dessous se trouve un arbre décisionnel pour le type de régression à employer avec les données de comptage. Il est toutefois important de mentionner que se fier à un test pour choisir une régression plutôt qu'une autre ne nous permet pas d'affirmer la non-validité de l'option jugée moins appropriée, mais seulement de dire que l'autre est plus adéquate. Pour fins de robustesse, nous gardons donc toutes les régressions effectuées.

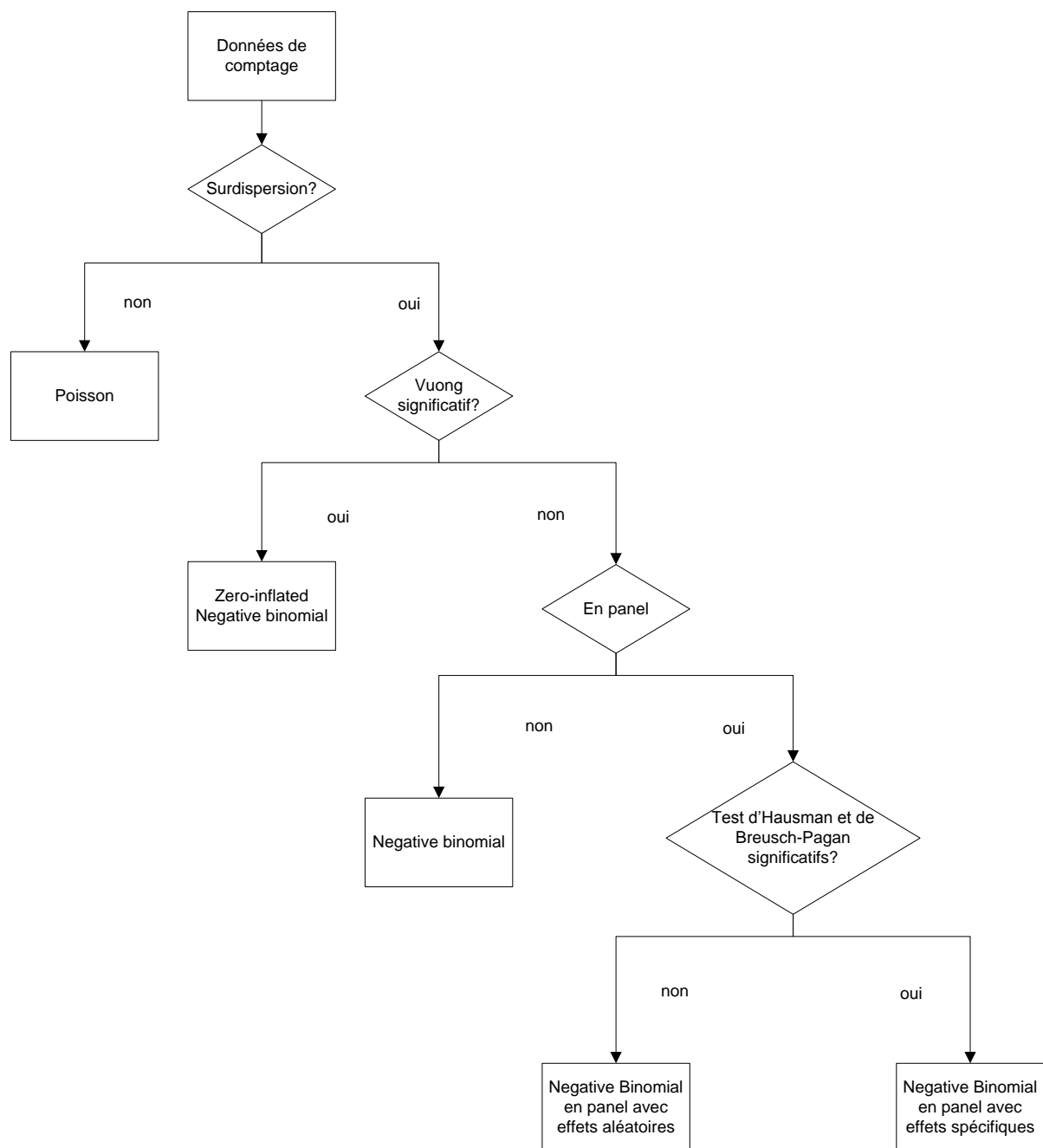


Figure 3-7: Arbre décisionnel des régressions pour les données de comptage.

### 3.4.8.4 Endogénéité

L'endogénéité, c'est-à-dire la corrélation entre un des termes explicatifs et l'erreur  $\mu$ , est non désirable dans l'application de méthodes économétriques car elle est source de biais. Cela est causé soit par (1) un oubli ou la non-disponibilité d'une variable explicative qui est corrélée à une autre variable indépendante, (2) une erreur de mesure qui se trouve alors dans l'erreur  $\mu$  ou (3) lorsqu'il y a simultanée. Nous faisons de notre mieux pour réduire le risque d'endogénéité en incluant le plus de variables pertinentes dans nos régressions pour éviter des problèmes liés aux causes (1) et (2). Par contre le cas (3) est prévalent surtout parce que divers phénomènes non observables expliquent à la fois nos variables dépendantes et certaines des variables endogènes.

Cette endogénéité causée par simultanée est illustrée de la manière suivante :

Ayant

$$y_1 = \alpha_1 y_2 + x_1 \beta_1 + \mu_1$$

Et

$$y_2 = x_2 \beta_2 + \mu_2$$

Une corrélation entre les erreurs  $\mu_1$  et  $\mu_2$  implique une corrélation entre  $y_2$  et  $\mu_1$ , ce qui est à éviter.

Pour tester s'il y a endogénéité et ainsi évaluer si un modèle à variables instrumentales ne serait pas une meilleure option nous employons le test de Durbin-Wu-Hausman<sup>35</sup> ou la méthode *two-stage residual inclusion*<sup>36</sup> (2SRI) (Terza et al., 2008). Ceci consiste à :

- (1) Effectuer une régression de la variable considérée endogène sur ses variables instrumentales et toutes les autres variables exogènes.
- (2) Sauvegarder les résidus de cette régression.

---

<sup>35</sup> Il existe plus d'un moyen de vérifier l'endogénéité, Stata possède même des commandes intégrés, tels *ivendog*, mais cette méthode est éprouvée et suffisante.

<sup>36</sup> Inclusion des résidus à deux étapes.

- (3) Effectuer la régression de la variable dépendante sur la variable endogène, les résidus précédemment sauvegardés et toutes les variables exogènes, excluant les variables instrumentales.

Des résidus non-significatifs indiquent qu'il n'y a pas d'endogénéité pour les variables spécifiées et qu'il n'est donc pas nécessaire de modéliser avec les variables instrumentales.<sup>37</sup>

Le 2SRI est idéal pour vérifier l'endogénéité avec des régressions à binomiale négative et peut même être utilisée en binomiale négative avec panel. Cependant, certains types de modèles ne supportant pas le 2SRI, il devient nécessaire de traiter l'endogénéité autrement. En particulier, cela est le cas pour la régression à binomiale négative à zéro-augmenté. La procédure à suivre devient alors un double moindres carrés ou *two-stage least squares* (2SLS) (Stephan et al., 2007) qui est assez similaire au 2SRI et qui permet de briser le lien entre variable dépendante et l'erreur  $\mu$ . Elle consiste à :

- (4) Effectuer une régression de la variable considérée endogène sur ses variables instrumentales et toutes les autres variables exogènes.
- (5) Sauvegarder l'estimation de la variable endogène prédite de cette régression.
- (6) Effectuer la régression de la variable dépendante sur la variable endogène prédite et toutes les variables exogènes, excluant les variables instrumentales.

Obtenir une variable endogène prédite significative indique qu'il y a endogénéité et qu'il est préférable de se fier à un modèle qui considère les variables instrumentales.

La modélisation de l'endogénéité et la sélection des variables endogènes et instrumentales se fait en prenant certaines considérations. Premièrement, les variables instrumentales doivent être en mesure de bien expliquer la variable endogène. Deuxièmement, par définition les variables instrumentales ne doivent pas être corrélées entre elles, ni avec la variable dépendante.

L'endogénéité est testée sur trois variables principales : les articles, les subventions et les contrats. Elles sont considérées endogènes entre autres parce qu'elles sont responsables pour

---

<sup>37</sup> Le même principe s'applique lors de régressions en panel, sauf qu'il faut tenir compte des trois types de résidus,  $u$ ,  $e$ , et  $ue$  en les isolant une à la fois dans la procédure du 2SRI.

expliquer les variables dépendantes, mais aussi parce qu'elles peuvent être considérées dans certaines instances dépendantes de certaines de ces variables. Par exemple, les contrats peuvent tant bien expliquer le nombre de brevets sur lequel un chercheur a participé que le nombre de brevets peut expliquer les contrats futurs reçus. Ci-dessous se trouvent les variables endogènes accompagnées de leurs variables instrumentales.

$$[nbarticlemoyan3] = f \left\{ \begin{array}{l} ltotsubvmoyan3, sqltotsubvmoyan3, ldmbetcent3, \\ codechair33, age, sq\_age, nbarticlemoyan3wuu \end{array} \right\}$$

$$[ltotsubvmoyan3] = f \left\{ \begin{array}{l} nbarticlemoyan3, ldmbetcent3, codechair33, \\ age, sq\_age, ltotsubvmoyan3wuu \end{array} \right\}$$

$$[ltotcontmoyan3] = f \left\{ \begin{array}{l} nbloop\_past, totcontpast, codechair33, \\ age, sq\_age, ltotcontmoyan3wuu \end{array} \right\}$$

### 3.4.8.5 Modèle Tobit

Le modèle de Tobit est employé pour la régression sur une variable dépendante tronquée. Cette méthode nous est donc utile seulement si nous rencontrons un skewness important. Nous convertissons la variable de comptage en logarithme naturel et sa valeur limite inférieure devient alors 0. Nous utilisons  $\ln(y+1)$  lorsque  $y \in (0, \infty)$  car  $\ln(0)$  donne  $-\infty$ .

Lorsque nous détectons de l'endogénéité avec la tobit, il devient nécessaire de modéliser la tobit à variables instrumentales (ivtobit).

### 3.4.8.6 Évaluation de la qualité de l'ajustement

Dans cette section, nous abordons la construction des modèles : c'est-à-dire comment les variables et les modèles ont été sélectionnés et comment l'endogénéité a été modélisée. La vérification des distributions des variables dépendantes et la compilation des matrices de corrélations sont aussi au sujet. Nous créons plusieurs modèles dans l'analyse de données. Ces

modèles varient en fonction des variables endogènes et exogènes et aussi selon l'objectif de l'analyse.

Une méthode qui permet de comparer la qualité des régressions est le coefficient de détermination, c'est-à-dire le  $R^2$ . Le  $R^2$  indique la capacité d'une régression à expliquer la variance dans le modèle. Il se calcule ainsi :

$$R^2 = 1 - \frac{SSE}{SST}$$

Ayant,

$$SST = SSR + SSE$$

Où

Somme des carrés totaux : SST

Somme des carrés de régression : SSR

Somme des carrés des erreurs : SSE

Cependant ce critère n'est utile que lorsque nous comparons des modèles ayant le même nombre de variables car sa valeur croît en fonction du nombre de variables sélectionnées. Sa valeur ajustée, i.e. tenant compte du nombre de variables, proposée par Theil (1971) est donc un meilleur outil lorsque généré. Son équation se trouve dans Judge et al. (1985, page 863)

Une autre méthode que nous utilisons pour déterminer si une variable est apte à être utilisée comme instrument ou exogène est la p-value des coefficients  $\beta$ .

Considérant que nous avons des effets temporels à modéliser, nous employons la fonction « L. » de Stata qui permet d'obtenir la valeur passée d'une variable quelconque. « L. » indique une variable au temps t -1, « L2. » représente un laps de t-2. Ceci est utile surtout pour la modélisation du financement. Par exemple, si nous voulons évaluer si un chercheur i a déposé un ou plusieurs brevets [nbpat] au temps t, nous analyserons les contrats octroyés dans les années précédant le temps t, à t-1 avec [L.ltotcontmoyan3] ou à t-2 avec [L2.ltotcontmoyan3]. Un autre exemple, même plus pertinent, serait dans un cas où nous modélisons un 2SRI, lorsque vient le



temps de faire la première régression; les instruments devant être représentatifs de la variable supposément endogène, certaines devront être déphasées d'une année supplémentaire par rapport à la variable dépendante, i.e dans le cas où nous avons [nbpat] comme variable dépendante, [L.ltotsubvmoyan3] comme variable endogène et [L2.ldmbetcent3] comme instrument de [L.ltotsubvmoyan3].

Afin de déterminer quels modèles retenir, nous nous fions à l'Akaike information Criterion (AIC) et au Bayesian Information Criterion (BIC)<sup>38</sup>. Ces critères d'information permettent de consolider à la fois la demande de précision d'estimation et de la meilleure approximation à la réalité (Judge et al., 1985). Ainsi, ils sont des moyens de classer les modèles entre eux et éventuellement sélectionner les meilleurs.

L'équation de l'AIC est la suivante :

$$AIC = -\frac{2}{T} \ln \ell(b_1^*|y) + \frac{2K_1}{T}$$

Où

$\ell(b_1^*|y)$  est la vraisemblance maximale et  $K_1$  est le nombre de paramètres

L'équation du BIC est donnée par :

$$BIC = \ln y' M_1 y + 2 \left[ (K_1 + 2) \frac{\hat{w}^2}{\hat{\sigma}^2} - \left( \frac{\hat{w}^2}{\hat{\sigma}^2} \right)^2 \right]$$

Où

$\hat{w}^2$  est un estimé indépendant de la vraie variance

$y \sim N(X\beta_0, \sigma_0^2 I)$

L'idée est de conserver les modèles avec les AIC et BIC minimaux. La procédure que nous suivons est empruntée de Burnham & Anderson (2002) : Essentiellement, nous pouvons totalement rejeter les modèles qui se trouvent à être plus grands que  $AIC_{\min}$  et  $BIC_{\min}$  par plus de 10, car ils expliquent moins bien la variation structurelle des données. Ceux qui se trouvent entre

---

<sup>38</sup> Aussi connue sous le nom de *Sawa's Criterion*

4 et 7 des minimums présentent un peu plus de pertinence, mais toujours relativement négligeable. Ceux dont les critères sont inférieurs à 3 par rapport au minimum peuvent être considérés comme des alternatives et ne sont donc pas à rejeter automatiquement.

La vérification de la corrélation entre les variables est aussi essentielle pour la création de modèles valides. La construction de la matrice de corrélation entre paires permet entre autres de détecter une interaction entre deux variables explicatives. Ceci est important dans la sélection de variables. En effet, il se peut qu'une variable  $x$  seule n'ait pas d'effet sur un modèle mais qu'en combinaison avec une autre possède une interaction significative. Cela est le souvent le cas avec les variables d'âge et de densité égocentrique par exemple.

Cela aide aussi à confirmer le rôle de certaines variables. Par exemple, nous ne pouvons utiliser une variable endogène si cette dernière se trouve à être corrélée avec la variable dépendante. La construction des matrices de corrélation est aussi importante lors de la vérification de l'endogénéité. Il est important de ne pas inclure des variables instrumentales qui sont fortement corrélées dans la même régression de première étape du test de Durbin-Wu-Hausman, car cela pourrait engendrer un biais.

La fonction `pwcorr` de Stata permet de générer les corrélations en paires et leur échelle va de -1 à 1. la limite que nous utilisons pour indiquer qu'une corrélation est significative est  $(-)0,5$ . Ci-dessous se trouvent les matrices de corrélations effectuées lors des régressions sur le nombre de revendications et considérant les subventions comme variable endogène. Les matrices de corrélation générales se trouvent à l'annexe 13.

Tableau 3-2 : matrice de corrélation entre la variable endogène et les variables instrumentales de [L.ltotsubvmoyan3]

	L.ltotsubvmoyan3	L2.nbarticlemoyan3	L2.ldmbetcentind3	L.ltotsubvmoyan3wuu
L.ltotsubvmoyan3	1.0000 9062			
L2.nbarticlemoyan3	0.1483 9062	1.0000 9068		
L2.ldmbetcentind3	0.1039 8789	0.7963 8795	1.0000 8795	
L.ltotsubvmoyan3wuu	0.3204 9045	0.0570 9051	-0.0404 8778	1.0000 9051

Tableau 3-3 : matrice de corrélation entre la variable dépendante et les variables instrumentales

	Inbclaims	L2.nbarticlemoyan3	L2.ldmbetcentind3	L.ltotsubvmoyan3wu
Inbclaims	1.0000 534			
L2.nbarticlemoyan3	0.0362 534	1.0000 9068		
L2.ldmbetcentind3	0.0525 526	0.7963 8795	1.0000 8795	
L.ltotsubvmoyan3wu	-0.0265 534	0.0570 9051	-0.0404 8778	1.0000 9051

Tel que mentionné plus haut, un des avantages à générer les tables de corrélation entre les variables est le fait que cela nous permet d'identifier des liens indésirables entre les variable. Dans la table 3-2 nous voyons qu'il y a une corrélation entre deux des variables instrumentales, soient [L2.nbarticlemoyan3] et [L2.ldmbetcentind3], c'est pourquoi nous devons les séparer lors de la modélisation de l'endogénéité sur [L.ltotsubvmoyan3].

### 3.4.8.7 Les Modèles

L'analyse économétrique a deux buts principaux : (1) le premier vise à déterminer les caractéristiques, attributs et circonstances qui font qu'un chercheur va breveter combiné à la qualité de ses brevets en nanotechnologie et en biotechnologie; (2) le deuxième vise à comprendre exclusivement ce qui justifie la qualité des brevets obtenus, i.e. pour voir si ce sont les mêmes éléments qui expliquent à la fois les citations et les revendications. Ici nous faisons la mise au point des modèles générées, modèles qui ont été créés en cascade selon les procédures énoncées plus haut.

(1) Le premier type de modèle est effectué sur un grand échantillon qui inclut plus de 8000 instances de chercheurs-années. Nous créons trois blocs de régressions pour modéliser le brevetage du chercheur et la qualité de ces brevets

Le **bloc A** combine le fait de breveter au nombre de brevets obtenus à l'année en cours et sa variable dépendante est [nbpat]. Nous avons affaire à des données de comptage et les modèles

n'étant pas trop sévères quant à la distribution de la variable dépendante, les régressions compilées sont la régression de Poisson<sup>39</sup>, la régression à binomiale négative (nbreg), à binomiale négative en panel (xtnbreg) car il y a assez d'observations en moyenne par groupe<sup>40</sup>, à binomiale négative à zéro-augmenté (zinb) et la tobit. En se fiant à la littérature et aux hypothèses énoncées, nous modélisons [nbpat] de la manière suivante :

$$[nbpat] = f \left\{ \begin{array}{l} \text{nbarticlemoyan3, sq\_nbarticlemoyan3, totcontmoyan3,} \\ \text{sqtotcontmoyan3, totsubvmoyan3, sqtotsubvmoyan3,} \\ \text{lmcliqind23, sq\_lmcliqind23, deg\_app0, avg\_pat\_past,} \\ \text{o\_pat\_past, years, clusters, nano\_excl, bio\_excl} \end{array} \right\}$$

Le **bloc B** est similaire au bloc A sauf qu'au lieu de considérer [nbpat] comme variable dépendante, nous considérons le nombre de revendications associées aux brevets d'un chercheur à une année spécifique [nbclaims]. Parce que nous avons des données de comptage, les régressions générées sont aussi à binomiale négative, à binomiale négative en panel, à binomiale négative à zéro-augmenté et la tobit. Le nombre de revendications est expliqué ainsi :

$$[nbclaims] = f \left\{ \begin{array}{l} \text{nbarticlemoyan3, sqnbarticlemoyan3, totcontmoyan3,} \\ \text{sqtotcontmoyan3, totsubvmoyan3, sqtotsubvmoyan3,} \\ \text{lmcliqind23, sq\_lmcliqind23, deg\_app0,} \\ \text{avg\_pat\_past, years, clusters, nano\_excl, bio\_excl} \end{array} \right\}$$

Le **bloc C** est identique au bloc B, sauf que la variable dépendante employée est [nbcitp5], i.e. le nombre de citations reçues après une période de 5 ans. La raison pour laquelle les blocs B et C contiennent exactement les mêmes régressions et variables explicatives est pour maintenir une congruence dans l'analyse subséquente.

(2) Le deuxième type de modèle est celui qui tente d'expliquer la qualité de brevets. Ici nous effectuons une transition pour ne considérer que les années auxquelles un chercheur a un ou plusieurs brevets. Cet échantillon est donc restreint à quelques 420 observations.

---

<sup>39</sup> Il est à mentionné cependant que la régression de Poisson a été omise complètement tant dans le bloc A que dans les blocs suivants, car il y a une présence très importante de surdispersion.

<sup>40</sup> 6 ans en moyenne par individu.

Le **bloc D** examine le nombre d'inventions brevetées auxquelles un chercheur a contribué selon l'année. Il est utile pour tester la robustesse du bloc A, mais aussi pour faire le passage à l'évaluation de la qualité, pour les instances où les chercheurs ont au moins un brevet. La variable dépendante est [nbpat] et les régressions possibles sont naturellement la nbreg et la tobit. Le modèle à estimer ici est :

$$[nbpat] = f \left\{ \begin{array}{l} \text{nbarticlemoyan3, sqnbarticlemoyan3, totcontmoyan3,} \\ \text{sqtotcontmoyan3, totsubvmoyan3, sqtotsubvmoyan3,} \\ \text{lmcliqind23, sq\_lmcliqind23, deg\_app0, avg\_pat\_past,} \\ \text{o\_patpast, years, nbloop0}_{t-1}, \text{clusters, nano\_excl, bio\_excl} \end{array} \right\}$$

Le **bloc E** contient les régressions sur la variable [nbclaims]. Celle-ci étant une variable discrète, il est possible de procéder avec une régression à binomiale négative. Il n'est cependant pas possible d'estimer une régression à binomiale négative en panel parce qu'il n'y a pas suffisamment d'observations par groupe (1,9 en moyenne). La binomiale négative à zéro-augmenté et la tobit sont aussi hors de question, car il n'y a pas d'instance où [nbclaims] vaut zéro dans l'échantillon. La seule autre régression qu'il est possible d'effectuer est celle des moindres carrés ordinaires en employant de logarithme naturel de [nbclaims], ce que nous faisons. Voici le modèle :

$$[nbclaims] = \left\{ \begin{array}{l} \text{nbarticlemoyan3, totcontmoyan3, sqtotcontmoyan3,} \\ \text{totsubvmoyan3, sqtotsubvmoyan3, lmcliqind23,} \\ \text{sq\_lmcliqind23, deg\_app, codechair33, age, sq\_age, nb\_loop,} \\ \text{propuninv, years, clusters, excl\_nano\_pat, excl\_bio\_pat} \end{array} \right\}$$

Le **bloc F** évalue le nombre de citations reçues sur 5 ans [nbcitp5]. Avec l'échantillon à l'étude, la distribution de cette variable permet l'utilisation des régressions à binomiale négative qui seront comparées aux résultats de binomiale négative du bloc E. Nous employons aussi la régression de tobit pour laquelle les conditions d'utilisation sont remplies.

Voici dans le tableau 3-5 ci-dessous, la liste abrégée des régressions finales réalisées dans le cadre de ce travail.

Tableau 3-4: Liste des régressions finales effectuées et des procédures de Stata utilisées pour chacun des modèles

Bloc de régressions Variable dépendante	Phase 1 (N ~ 8000)			Phase 2 (N ~ 420)		
	A nbpat	B nbclaims	C nbcitp5	D nbpat	E nbclaims	F nbcitp5
	nbg	nbg	nbg	nbg	nbg	nbg
	xtnbg	xtnbg	xtnbg	tobit	reg	tobit
	zinb	zinb	zinb			
	tobit	tobit	tobit			

Il est important de noter que nous testons d'ailleurs l'endogénéité sur trois variables [nbarticlemoyan3], [ltotsubvmoyan3] et [ltotcontmoyan3] pour chacune des régressions contenues dans chacun des blocs. Ainsi pour les régressions par moindres carrés ordinaire et à binomiale négative (avec et sans panel), nous effectuons du 2SRI, pour la régression à binomiale négative à zéro-augmenté nous estimons un modèle 2SLS et pour la tobit nous générons la tobit à variables instrumentales (ivtobit).

Tableau 3-5 : Régressions du bloc A sur le fait de breveter combiné au nombre de brevets

Bloc:	A
Version:	7
Modèles:	NBREG (2SRI, 2SLS), XTNBREG (2SRI, 2SLS), ZINB (2SLS), TOBIT(IV)
Variable dépendante:	nbpat, lnbpap

[illegible]

Tableau 3-6: Régressions du bloc B sur le nombre de revendications combinant le fait de breveter au nombre comme tel

Bloc:	B
Version:	7
Modèles:	NBREG (2SRI, 2SLS), XTNBREG (2SRI, 2SLS), ZINB (2SLS), TOBIT(IV)
Variable dépendante:	nbclaims, lnclaims

[illegible]





Tableau 3-8 : Régressions du bloc D sur le nombre de brevets si et seulement si le chercheur possède au moins un brevet

Bloc: D  
Version: 7  
Modèles: NBREG (2SRI, 2SLS), TOBIT(IV)  
Variable dépendante: nbpat, lnbp

[illegible]

Tableau 3-9: Régressions du bloc E sur le nombre de revendications si et seulement si le chercheur possède au moins un brevet

Bloc: E  
 Version: 7A  
 Modèles: REG (2SRI, 2SLS), NBREG (2SRI, 2SLS)  
 Variable dépendante: nbclaims, lnclaims

Variables	sans	avec modélisation de l'endogénéité sur									
	endogénéité	L.nbarticlemoyan3		L.ltotsubvmoyan3		L.ltotsubvmoyan3		L.ltotsubvmoyan3		L.totcontmoyan3	
	indépendantes	exogènes	instruments	exogènes	instruments	exogènes	instruments	exogènes	instruments	exogènes	instruments
L.nbarticlemoyan3	x										
L2.nbarticlemoyan3					x			x			
L2.sq_nbarticlemoyan3								x			
L.totcontmoyan3	x	x		x		x		x			
L.sq_ltotcontmoyan3	x	x		x		x		x			
L.lmcliqind23	x	x		x		x		x			x
L.sq_lmcliqind23	x	x		x		x		x			x
deg_app	x	x		x		x		x			x
years	x	x		x		x		x			x
clusters	x	x		x		x		x			x
L.ltotsubvmoyan3											x
L.sq_ltotsubvmoyan3											x
L2.ltotsubvmoyan3			x								
L2.sq_ltotsubvmoyan3			x								
L2.lmbetcent3			x		x				x		
codechair33	x	x		x		x		x			
L.age	x	x		x		x		x			
L.sq_age	x	x		x		x		x			
L.nbarticlemoyan3wuu			x								
L.totsubvmoyan3wuu					x			x			
L.totcontmoyan3wuu									x		
L.nb_loop_past											x
L.totcont_past											x
nb_loop	x	x		x		x		x		x	
propuninv	x	x		x		x		x		x	
excl_nano_pat	x	x		x		x		x		x	
excl_bio_pat	x	x		x		x		x		x	

Tableau 3-10: Régressions du bloc F sur le nombre de citations si et seulement si le chercheur possède au moins un brevet

Bloc: F  
 Version: 7A  
 Modèles: NBREG (2SRI, 2SLS), TOBIT(IV)  
 Variable dépendante: nbcitp, lnbcitp5

Variables	sans	avec modélisation de l'endogénéité sur							
	endogénéité	L.nbarticlemoyan3		L.ltotsubvmoyan3		L.ltotsubvmoyan3		L.ltotsubvmoyan3	
	indépendantes	exogènes	instruments	exogènes	instruments	exogènes	instruments	exogènes	instruments
L.nbarticlemoyan3	x								
L2.nbarticlemoyan3					x		x		
L2.sq_nbarticlemoyan3							x		
L.totcontmoyan3	x	x		x		x		x	
L.sq_ltotcontmoyan3	x	x		x		x		x	
L.lmcliqind23	x	x		x		x		x	x
L.sq_lmcliqind23	x	x		x		x		x	x
deg_app	x	x		x		x		x	x
years	x	x		x		x		x	x
clusters	x	x		x		x		x	x
L.ltotsubvmoyan3									x
L.sq_ltotsubvmoyan3									x
L2.ltotsubvmoyan3			x						
L2.sq_ltotsubvmoyan3			x						
L2.ldmbetcent3			x		x			x	
codechair33	x	x		x		x		x	
L.age	x	x		x		x		x	
L.sq_age	x	x		x		x		x	
L.nbarticlemoyan3wuu			x						
L.totsubvmoyan3wuu					x		x		x
L.totcontmoyan3wuu									x
L.nb_loop_past									x
L.totcont_past									x
nb_loop	x	x		x		x		x	x
propuninv	x	x		x		x		x	x
excl_nano_pat	x	x		x		x		x	x
excl_bio_pat	x	x		x		x		x	x

## **CHAPITRE 4    RÉSULTATS ET ANALYSE**

Dans ce chapitre nous explorons d'abord les similarités entre la nanotechnologie et la biotechnologie, l'importance du financement pour le brevetage, l'implication des firmes et des scientifiques en nanotechnologie comme en biotechnologie et la nature de ces collaborations. Nous passons ensuite à l'analyse des régressions.

### **4.1 Similarité en nanotechnologie et biotechnologie**

Prendre les diverses mesures de qualité et le degré d'application de connaissances associées à des inventions brevetées nous permet de mener une analyse de similarité (comparaison de moyenne) originale entre les brevets canadiens en nanotechnologie et en biotechnologie. Avant de porter notre regard sur les boucles d'innovation comme telles, examinons d'abord les caractéristiques des innovations brevetées de manière générale. Afin de limiter le biais engendré par des effets de taille, nous divisons l'échantillon de biotechnologie en 3 sous-échantillons pour les comparer à la nanotechnologie (ensemble A) et en 4 sous-échantillons pour les comparer à la nanobiotechnologie (ensemble B). Le Tableau 4-1 ci-dessous présente les statistiques de base et le Tableau 4-2 suivant illustre les résultats des tests de comparaison. En ce qui concerne la qualité, notons que les brevets de nanotechnologie présentent une supériorité du nombre de revendications et de citations après 5 ans statistiquement significative. Ceci suggère qu'ils ont une plus haute probabilité de faciliter des innovations futures et sont utilisés de manière plus fréquente pour générer de nouvelles inventions brevetées, i.e. qu'ils sont possiblement de meilleure qualité.

Considérant le degré d'application des connaissances associées à tous les brevets canadiens, nous trouvons que les connaissances publiées associées à la nanotechnologie sont plus appliquées que celles associées à la biotechnologie. Il faut mentionner par contre que même si ces moyennes sont statistiquement significatives, nous considérons la différence actuelle entre les deux catégories comme minime car l'échelle du degré d'application des journaux est discrète, allant de 1 à 4.

Tableau 4-1 : Qualité des brevets et degré d'application

		NANOTECHNOLOGIE		BIOTECHNOLOGIE				NANOBIOTECHNOLOGIE	
				A		B			
		Moyenne (Dév. Std.)	n	Moyenne (Dév. Std.)	n	Moyenne (Dév. Std.)	n	Moyenne (Dév. Std.)	n
Moyenne du nombre de revendications	Sample No								
	1	21.49 (15.09)	3555	17.16 (14.70)	3457	17.13 (13.94)	2592	21.06 (16.24)	2386
	2	-	-	17.44 (14.79)	3477	17.00 (14.58)	2625	-	-
	3	-	-	16.94 (13.59)	3448	17.48 (14.57)	2603	-	-
	4			-	-	17.12 (14.38)	2562	-	-
	Total	21.49	3555	17.18	10382	17.18 (14.37)	10382	21.06	2386
Moyenne du nombre de citations après 5 ans	Sample No								
	1	3.96 (7.08)	3577	2.31 (3.93)	3476	2.53 (4.73)	2606	3.09 (6.11)	2404
	2	-	-	2.44 (4.34)	3496	2.29 (3.78)	2641	-	-
	3	-	-	2.36 (3.94)	3465	2.38 (3.94)	2615	-	-
	4			-	-	2.30 (3.79)	2575	-	-
	Total	3.96	3577	2.37	10437	2.37	10437	3.09	2404
Moyenne du degré d'application	Sample No								
	1	3.05 (0.68)	861	3.36 (0.69)	590	3.37 (0.72)	462	3.25(0.72)	802
	2	-	-	3.36 (0.67)	606	3.33 (0.72)	440	-	-
	3	-	-	3.37 (0.67)	592	3.42 (0.62)	446	-	-
	4			-	-	3.33 (0.65)	440	-	-
	Total	3.05	861	3.36	1788	3.36	1788	3.25	802

Tableau 4-2 : Degré de signification des comparaisons de qualité et du degré d'application pour H0: égalité des moyennes.

		NANOTECH vs BIOTECH (A)	NANOTECH vs NANOBIOTECH	BIOTECH (B) vs NANOBIOTECH
Moyenne du nombre de revendications	Sample No			
	1	****	**	****
	2	****	-	****
	3	****	-	****
	4	-	-	****
Moyenne du nombre de citations après 5 ans	Sample No			
	1	****	****	+
	2	****	-	+
	3	****	-	+
	4	-	-	*
Moyenne du degré d'application	Sample No			
	1	****	****	***
	2	****	-	*
	3	****	-	****
	4	-	-	*

+ for  $p > 0.10$ , \* for  $p < 0.10$ , \*\* for  $p < 0.05$ , \*\*\* for  $p < 0.01$ , \*\*\*\* for  $p < 0.001$ . L'échantillon de nanotechnologie exclut la nanobiotechnologie et l'échantillon de la biotechnologie exclut la nanotechnologie.

## 4.2 Processus de bouclage

En associant le financement de chercheurs universitaires avec les données de brevetage, nous sommes capables d'identifier 1518 instances où le financement a mené au brevetage en biotechnologie et 1026 cas en nanotechnologie. Le tableau ci-dessous présente les statistiques sommaires concernant ces boucles. En contraste avec les Tableaux 4-1 et 4-2, le Tableau 4-3 illustre les statistiques sommaires pour la nanotechnologie et la biotechnologie incluant la biotechnologie dans les échantillons. Par exemple 24 des 25 firmes impliquées dans des boucles en nanotechnologie contribuent à la recherche en nanobiotechnologie. Nous trouvons qu'en général, le délai entre le financement et l'application du brevet et par la suite l'octroi du brevet est plus court en nanobiotechnologie et est plus onéreux en termes de contrat ou subvention que la nanotechnologie ou la biotechnologie.

Tableau 4-3 : Statistiques sommaires des boucle d'innovation

		Nanotechnologie		Biotechnologie		Nanobiotechnologie	
Nombre de boucles potentielles		1026		1518		673	
Nombre d'organisations pourvoyeurs/cessionnaires		37		61		31	
		Canadien	Non-Canadien	Canadien	Non-Canadien	Canadien	Non-Canadien
	Nombre de firmes	25	4(1)*	42	7(3)	24	2(1)
	Nombre d'universités	5	0	5	0	5	0
	Nombre de fondations, associations, sociétés	2	0	3	1	2	0
	Nombre d'institutions de santé	0	0	1	0	0	0
	Nombre de sociétés d'état	1	0	2	0	1	0
Nombre de scientifiques inventeurs		62		151		44	
Nombre de brevets développés en boucle		157		291		109	
Lapse de temps moyen entre financement et dépôt de brevet		4.2 (4.9)**		4.3 (4.6)		3.5 (4.8)	
Lapse de temps moyen entre financement et octroi de brevet		7.5 (8.2)		7.7 (7.9)		6.9 (8.2)	
Nombre moyen de boucles par chercheur		16.5		10.1		15.3	
Nombre moyen de boucle par institution		27.7		24.9		21.7	
Nombre de fonds octroyés aux chercheurs		325		691		211	
Somme total des fonds octroyés aux chercheurs		19 159 855\$ <sup>++</sup>		34 415 515\$		14 865 294\$	
Valeur moyenne des fonds octroyés		58 953\$		49 805\$		70 451\$	

\* Le nombre entre parenthèses représente les firmes ayant des affiliations domestiques et internationales et desquelles un financement a été octroyé directement via une source internationale. \*\* Intervalle moyen entre la liste distincte de fonds et de brevets inclus dans les boucles. ++ Tous les fonds sont exprimés en \$ CAN constants de 2002.



Tableau 4-4 : Statistiques descriptives de boucles d'innovation en nanotechnologie selon les sous-domaines

		Nanobiotechnologie		metrologie & processus		dispositifs & électronique		matériaux & structures	
Nombre de boucles potentiels		673		42		50		284	
Nombre d'organisations finançant/brevetant		31		6		6		9	
		Canadien	Non-Canadien	Canadien	Non-Canadien	Canadien	Non-Canadien	Canadien	Non-Canadien
	Nombre de firmes	24	2(1)*	2	1	3	0	3	1
	Nombre d'établissements d'enseignement	5	0	3	0	2	0	3	0
	Nombre de fondations, associations, sociétés	2	0	0	0	0	0	1	0
	Nombre d'institutions de santé	0	0	0	0	0	0	0	0
	Nombre de corporations provinciales	1	0	0	0	1	0	1	0
Nombre de scientifiques inventeurs		44		6		8		16	
Nombre de brevets développés par boucle		109		8		8		38	
Intervalle moyen entre financement et application de brevet		3,5 (4,8)**		3,8(4,1)		4,7(3,3)		4(3,5)	
Intervalle moyen entre financement et octroi de brevet		6,9 (8,2)		6,4(6,6)		7,2(6,3)		7,2(6,4)	
Nombre moyen de boucles par chercheur		15,3		7,0		6,3		17,8	
Nombre moyen de boucles par institution		21,7		7,0		8,3		31,6	
Nombre de fonds octroyés aux chercheurs		211		35		29		121	
Somme de fonds octroyés aux chercheurs		14865294 \$ ++		3 141 502 \$		1 856 959 \$		7 433 259 \$	
Valeur moyenne des fonds octroyés		70 451 \$		89 757 \$		64 033 \$		61 432 \$	

\* Les nombres entre parenthèses représentent les firmes ayant des affiliations domestiques et internationales et desquelles un financement a été octroyé directement via une source internationale. \*\* Intervalle moyen entre la liste distincte de fonds et de brevets inclus dans les boucles. ++ Tous les fonds sont exprimés en \$CAN constants de 2002.

Tableau 4-5: Statistiques descriptives de boucles d'innovation en biotechnologie selon les sous-domaines<sup>41</sup>

		Nanobiotechnologie		AND/ARN		Protéines et molécules	
Nombre de boucles potentiels		673		154		267	
Nombre d'organisations finançant/brevetant		31		14		27	
		Canadien	Non-Canadien	Canadien	Non-Canadien	Canadien	Non-Canadien
	Nombre de firmes	24	2(1)*	7	2	16	5
	Nombre d'établissements d'enseignement	5	0	5	0	5	0
	Nombre de fondations, associations, sociétés	2	0	0	0	0	1
	Nombre d'institutions de santé	0	0	0	0	1(1)	0
	Nombre de corporations provinciales	1	0	0	0	0	0
Nombre de scientifiques inventeurs		44		29		50	
Nombre de brevets développés par boucle		109		39		60	
Intervalle moyen entre financement et application de brevet		3,5 (4,8)**		4,3(4,2)		5,6(4,0)	
Intervalle moyen entre financement et octroi de brevet		6,9 (8,2)		7,8 (7,7)		9(7,3)	
Nombre moyen de boucles par chercheur		15,3		5,3		5,3	
Nombre moyen de boucles par institution		21,7		11,0		9,9	
Nombre de fonds octroyés aux chercheurs		211		92		191	
Somme de fonds octroyés aux chercheurs		14 865 294\$ ++		4 806 581 \$		10 464 759 \$	
Valeur moyenne des fonds octroyés		70 451 \$		52 245 \$		54 789 \$	

\* Les nombres entre parenthèses représentent les firmes ayant des affiliations domestiques et internationales et desquelles un financement a été octroyé directement via une source internationale. \*\* Intervalle moyen entre la liste distincte de fonds et de brevets inclus dans les boucles. ++ Tous les fonds sont exprimés en \$CAN constants de 2002.

<sup>41</sup> Ce tableau a été coupé pour fin de lisibilité, veuillez consulter la suite sur la prochaine page.

Tableau 4-6 (Suite): Statistiques descriptives de boucles d'innovation en biotechnologie selon les sous-domaines

		Ingénierie tissulaire et cellulaire		Techniques de processus biotechnologiques		Vecteurs de gènes et d'ARN		Bioinformatique	
Nombre de boucles potentiels		142		183		33		76	
Nombre d'organisations finançant/brevetant		13		15		6		6	
		Canadien	Non-Canadien	Canadien	Non-Canadien	Canadien	Non-Canadien	Canadien	Non-Canadien
	Nombre de firmes	8	0	8	1	4	0	3	0
	Nombre d'établissements d'enseignement	4	0	4	0	2	0	2	0
	Nombre de fondations, associations, sociétés	1	0	1	0	0	0	0	0
	Nombre d'institutions de santé	0	0	0	0	0	0	0	0
	Nombre de corporations provinciales	0	0	1	0	0	0	1	0
Nombre de scientifiques inventeurs		32		29		9		9	
Nombre de brevets développés par boucle		35		30		9		10	
Intervalle moyen entre financement et application de brevet		4,3(4,5)		6,4(5,5)		1,4(1,7)		3(3,3)	
Intervalle moyen entre financement et octroi de brevet		7,1 (7,5)		10,1(8,4)		4,4(4,9)		6,5(6,7)	
Nombre moyen de boucles par chercheur		4,4		6,3		3,7		8,4	
Nombre moyen de boucles par institution		10,9		12,2		5,5		12,7	
Nombre de fonds octroyés aux chercheurs		108		160		26		66	
Somme de fonds octroyés aux chercheurs		4 271 657 \$		7 071 474 \$		1 962 789 \$		2 864 544 \$	
Valeur moyenne des fonds octroyés		39 552 \$		44 197 \$		75 492 \$		43 402 \$	

\* Les nombres entre parenthèses représentent les firmes ayant des affiliations domestiques et internationales et desquelles un financement a été octroyé directement via une source internationale. \*\* Intervalle moyen entre la liste distincte de fonds et de brevets inclus dans les boucles. ++ Tous les fonds sont exprimés en \$CAN constants de 2002.

Afin de déterminer si le financement de la recherche universitaire est aussi important pour l'innovation privée qu'il est pensé être, nous commençons par examiner la contribution au brevetage des chercheurs universitaires financés au brevetage. Pour la nanotechnologie, la proportion d'inventions brevetées qui ont été produites par au moins un chercheur universitaire ayant reçu des fonds du cessionnaire est quelque peu limitée. La figure 4-1 montre que le financement privé à l'intérieur d'une boucle de chercheurs universitaires a mené à près de 1 % des brevets en nanotechnologie en 1991; ceci étant une contribution tangible du financement privé de la recherche universitaire. L'évolution du nombre de brevets résultant de financement en boucle d'innovation est montrée à la figure 4-1. L'implication de ces chercheurs suit une courbe en U inversée culminant à un peu moins de 6 % en 1999 et diminuant à 1 % en 2005 et stabilisant à 3 % en 2008. Les boucles en biotechnologie surviennent un peu plus tôt, en 1989, et suivent une courbe en S similaire avec un maximum de 5 % en 2006.

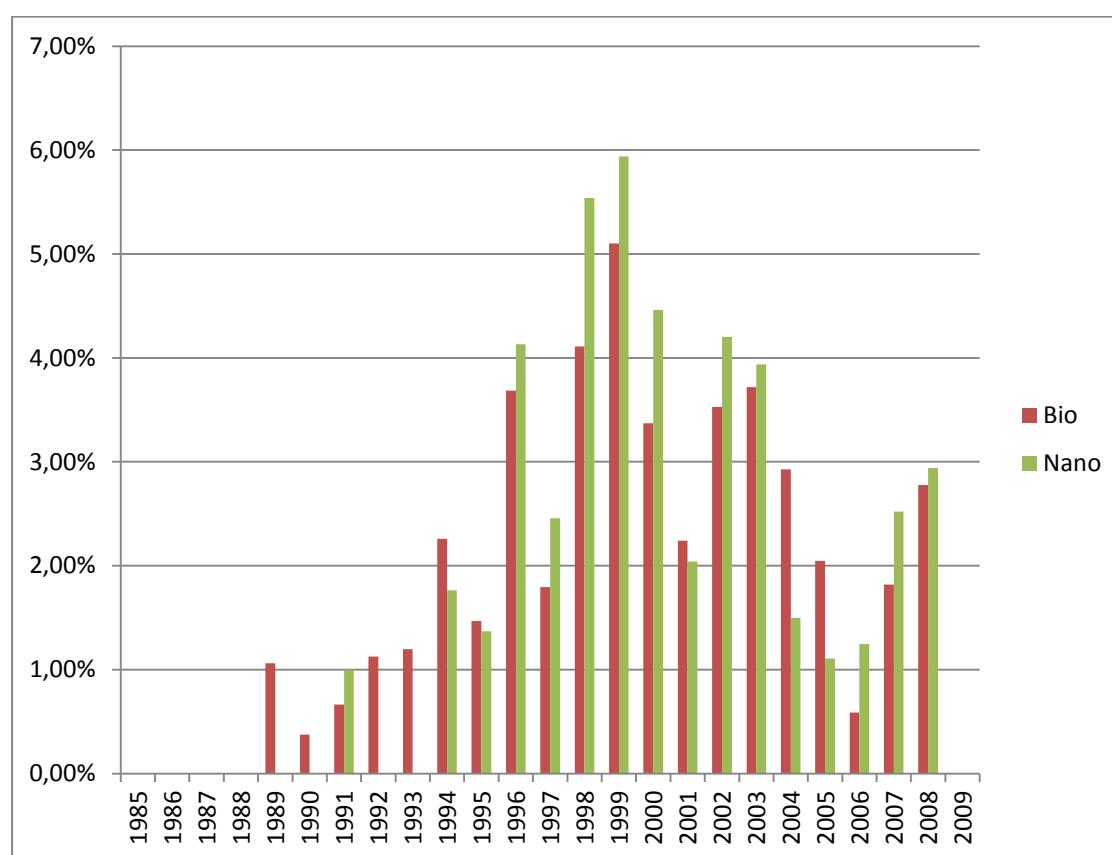


Figure 4-1: Proportion de brevets qui ont au moins un chercheur contenu dans une boucle

Ces résultats suggèrent qu'autant dans le cas de la nanotechnologie et la biotechnologie, étonnamment peu de chercheurs universitaires ont reçu du financement de firmes qui les mentionnent comme inventeurs sur des brevets qui sont la propriété d'une de ces organisations. L'importance de la collaboration université/industrie étant reconnue comme un facteur clef dans l'innovation en nanotechnologie et biotechnologie, deux domaines à bases scientifiques, nous pourrions nous attendre à une plus grande contribution d'inventeurs académiques, mais ceci n'est évidemment pas le cas, du moins au Québec.

Du total des 2545 cessionnaires de brevets biotechnologiques dans notre base de données complète, seulement 61 sont inclus dans un schéma de bouclage d'innovation. En d'autres mots, seulement 61 des cessionnaires de brevets biotechnologiques ont vu leur financement d'un chercheur universitaire québécois mener au brevetage d'une invention où ces inventeurs sont nommés sur le brevet. De ces 61 cessionnaires, 42 sont des firmes, 5 sont des institutions universitaires, 4 sont des fondations, 2 sont des sociétés d'états, et 1 est une institution hospitalière. Dans le cas de la nanotechnologie, il y a près de 1251 cessionnaires de brevets depuis le début des années 1980, mais seulement 37 de ceux-ci sont inclus dans des boucles dont 25 sont des firmes, 5 sont des institutions universitaires, 2 sont des fondations et 1 est une société d'état. Il est intéressant de remarquer que même s'il y a plusieurs milliers de firmes, fondations et institutions qui ont des droits sur les brevets de nanotechnologie et de biotechnologie, relativement peu ont financé la recherche universitaire qui a mené au brevetage. En moyenne, chaque organisation impliquée en nanotechnologie a participé dans 27,7 boucles, ce qui est légèrement plus que les 24,9 de ceux en biotechnologie (voir tableau 4-3). En fait, un petit nombre d'équipes cessionnaires/inventeurs-universitaires sont responsables d'une très grande majorité des boucles (figures 4-2 et 4-3) pendant qu'une majorité de ces équipes sont impliquées dans très peu de boucles. Ces partenariats illustrent parfaitement le principe de Pareto. Plusieurs des chercheurs universitaires qui se trouvent impliqués dans nos boucles semblent être des collaborateurs récurrents qui sont liés à plusieurs fonds et brevets ce que certains pourraient soutenir comme étant l'effet Saint-Matthieu<sup>42</sup> (Merton, 1968) jusqu'à une certaine mesure. Ceci

---

<sup>42</sup> Plus un individu a le potentiel de faire fructifier les ressources à sa disposition, plus il en recevra.

est à prévoir puisque le financement qui mène à l'innovation conduit souvent aussi à plus de financement (Van Looy et al., 2004). Les firmes préfèrent collaborer avec des chercheurs universitaires qui ont déjà fait une découverte reconnue ; ils donnent aussi plus aux scientifiques avec qui ils ont un historique de collaboration (Beaudry & Schiffauerova, 2011) jusqu'à un certain point cela facilite le progrès (Zucker & Darby, 1996). Qui plus est, les organisations impliquées en brevetage semblent aussi être impliquées dans les partenariats récurrents, c'est-à-dire qu'il y a peu d'organisations impliquées dans un grand nombre de boucles. Ceci est normal puisque la plupart des firmes sont seulement prêtes à financer les recherches d'universitaires s'ils ont la capacité de s'approprier les connaissances générées par ces chercheurs universitaires (Bonaccorsi & Piccaluga, 1994) et ceci n'est pas le cas pour la plupart des organisations.

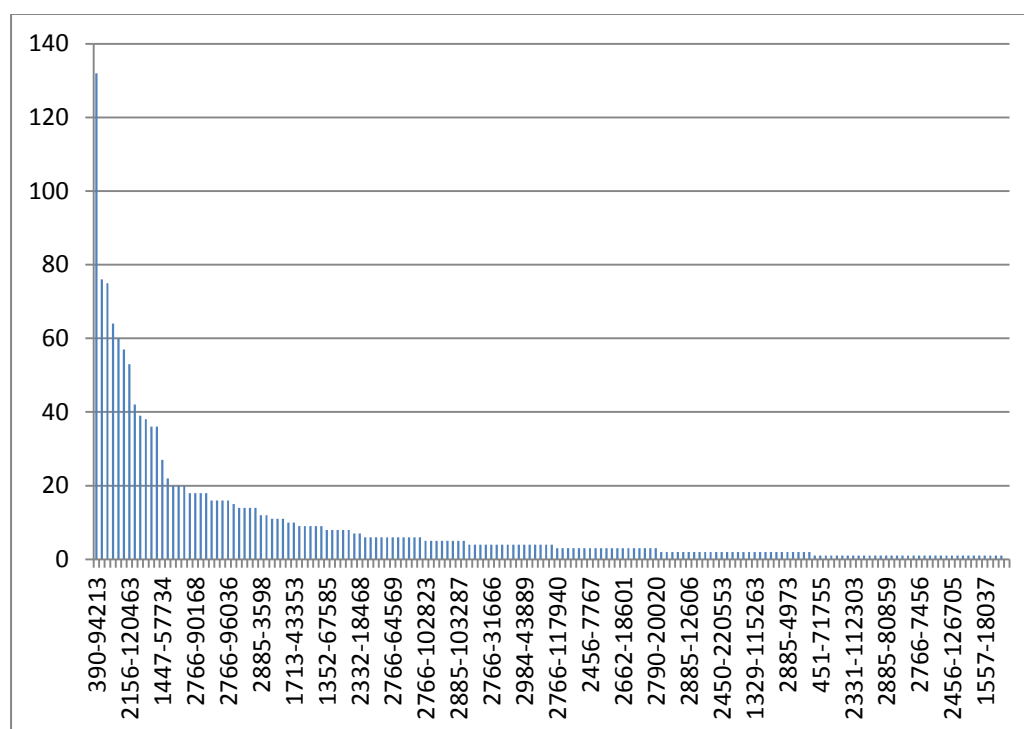


Figure 4-2: Charte de collaboration organisation-scientifique en biotechnologie (nombre de boucles d'innovation par partenariat)

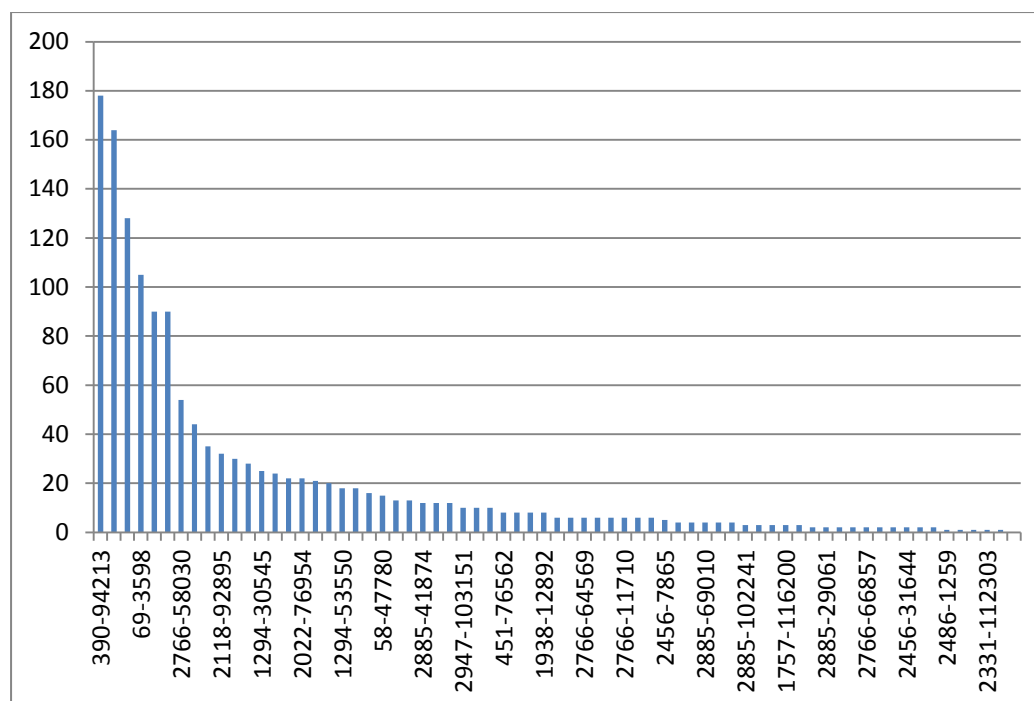


Figure 4-3: Charte de collaboration organisation-scientifique en nanotechnologie (nombre de boucles d'innovation par partenariat)

De plus, la moyenne des fonds menant à l'innovation est légèrement supérieure en nanotechnologie qu'en biotechnologie avec 58 953\$ et 49 805\$ respectivement. Ces moyennes cachent une croissance globale dans la valeur moyenne des fonds octroyés aux chercheurs universitaires qui mène au brevetage. Ceci peut être expliqué par le fait que, même si le nombre annuel de fonds menant au brevetage a diminué après 1999, la moyenne de financement reste constante après 2002.

La procédure de bouclage révèle que 151 scientifiques québécois en biotechnologie ont reçu du financement et sont mentionnés comme auteurs sur des brevets détenus par des organisations qui les ont financés. Dans le cas de la nanotechnologie, seulement 62 scientifiques sont de ce nombre, dont 44 sont en nanobiotechnologie. Comparé au montant actuel d'universitaires québécois qui sont responsables pour une invention brevetée soit en nanotechnologie ou en biotechnologie, ce nombre est faible. En fait, nous trouvons qu'il y a 786 chercheurs universitaires mentionnés comme inventeurs sur des brevets en biotechnologie, comparativement à 394 en nanotechnologie. La figure 4-4 illustre que les inventeurs universitaires contribuent à une grande portion des

brevets dans ces domaines scientifiques. En termes relatifs, ceci implique que 19,2 % des inventeurs-universitaires ont réellement reçu des fonds qui ont mené au brevetage avec l'organisation pourvoyeur et ce pourcentage diminue à 15,7% des inventeurs-universitaires en nanotechnologie.

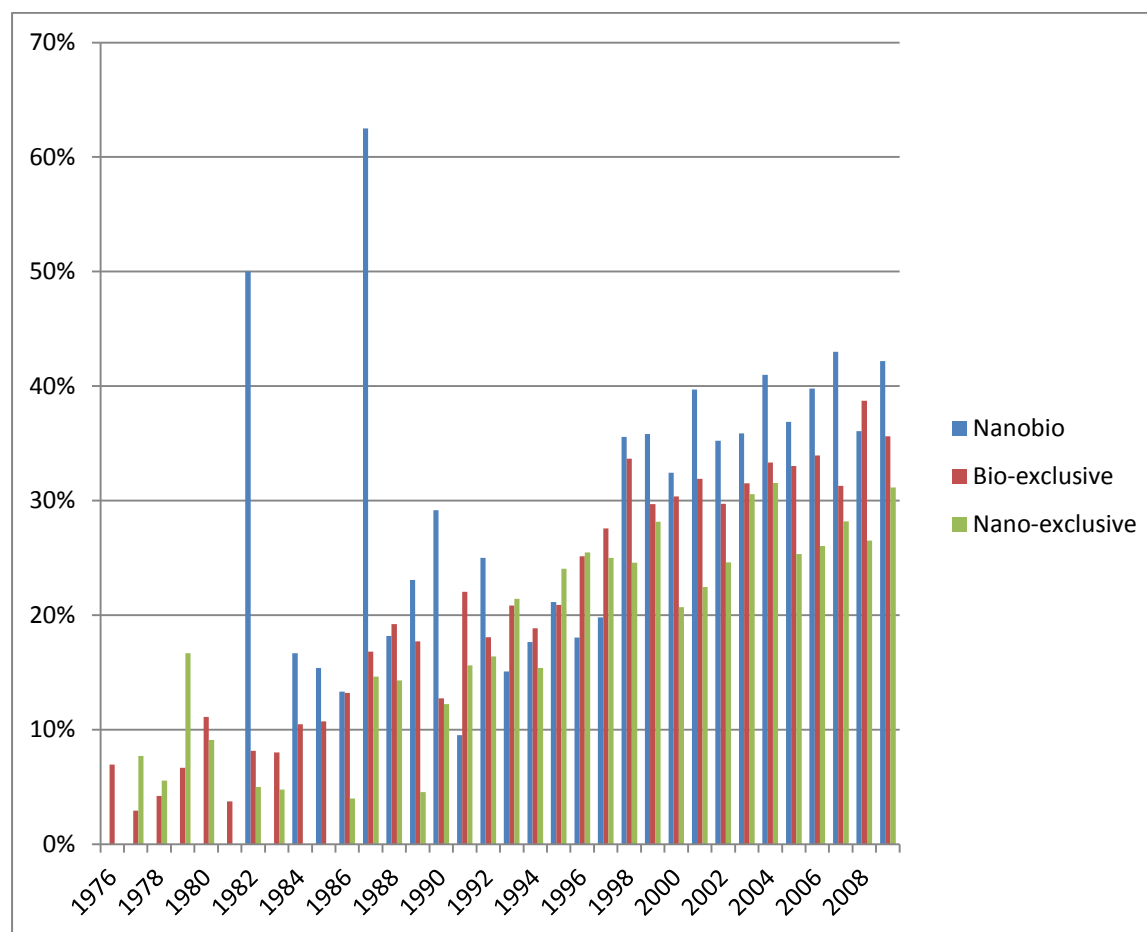


Figure 4-4: Proportion de brevets ayant des chercheurs universitaires comme inventeurs

En considérant, l'importance reconnue des chercheurs universitaires en innovation nanotechnologique et biotechnologique, nous nous attendions à une plus haute proportion. Il se peut cependant que les organisations ne considèrent pas les universitaires comme un moyen de breveter, i.e. la raison qui soutient le financement est liée à la recherche fondamentale qui ne mène pas nécessairement au brevet.



### 4.3 Qualité de brevet

Finalement nous examinons si les brevets inclus dans les boucles et ainsi qui sont issus de collaboration entre universitaires et firmes sont d'une plus grande qualité que ceux qui sont exclus de boucles. La partie supérieure du tableau 4-6 contient les moyennes du nombre de revendications des brevets appartenant à des firmes et reliées à des boucles d'innovation et les moyennes de celles hors-boucle. Il est très surprenant de voir qu'il n'y a pas de brevets exclusivement nanotechnologiques (excluant la nanobiotechnologie) qui sont la propriété de firmes québécoises et qui sont hors-boucles. Il est donc impossible de comparer le nombre de citations et le nombre de revendications avec ceux des brevets en biotechnologie. Il est possible toutefois de comparer les moyennes en biotechnologie.

Tableau 4-6 : Indicateurs de qualité des brevets appartenant à des firmes<sup>43</sup>

		NANOTECHNOLOGY		BIOTECHNOLOGY		NANOBIOTECHNOLOGY	
		Average (Std. Dev.)	n	Average (Std. Dev.)	n	Average (Std. Dev.)	n
Average Nb of claims	Category						
	In loop	24.04	24	18.0 (14.35)**	80	19.47 (17.58)+	109
	Out of loop	-	-	16.20 (17.65)**	297	15.93 (12.89)+	115
	Total	24.04	24	16.58	377	17.65	224
Average Nb of citations after 5 yrs	Category						
	In loop	2.46	24	1.56 (2.44)*	80	1.66 (2.99)+	109
	Out of loop	-	-	2.05 (3.52)*	298	1.84 (2.61)+	115
	Total	24.04	24	1.95	378	1.75	224

+p>0.10, \* for p<0.10, \*\* for p<0.05, \*\*\* for p<0.01, \*\*\*\* for p<0.001

Il semble que dans le cas de la biotechnologie, les brevets issus de collaboration financière entre chercheurs universitaires et firmes ont un plus haut taux de revendications que ceux exclus de cette collaboration financière directe.

Les résultats suggèrent aussi que les brevets en biotechnologie qui font partie de boucles impliquant de la recherche universitaire financée par l'industrie reçoivent systématiquement un plus bas taux de citations que ceux hors boucle. *C'est ce point-ci qui soulève un aspect*

---

<sup>43</sup> Pour cette analyse, nous limitons notre échantillon aux brevets étant la propriété de firmes québécoises afin de limiter le biais de firmes hors du Québec ayant financé des chercheurs non-québécois pour qui nous n'avons pas les données financières. Nous excluons aussi les brevets en boucle non-industriels.

*intéressant concernant les revendications et les citations comme moyens de mesure de qualité et qui est à l'origine de notre deuxième objectif: comparer les facteurs influençant et les différences entre les revendications et les citations.* Pour ce qui est de la nanobiotechnologie, les résultats sont similaires, mais malheureusement statistiquement non-significatifs. En prenant le nombre de revendications et le nombre de citations en considération, il est possible de suggérer que les brevets issus des boucles d'innovation desquels les cessionnaires sont des firmes québécoises ont un plus grand champ d'application que ceux qui sont hors boucles, mais qu'ils semblent être moins utiles pour faciliter de futures innovations (ou du moins celles qui sont brevetées). Si les citations sont considérées être la mesure dominante ou plus appropriée de la qualité, alors nous pouvons dire que les brevets en boucle sont de moindre qualité, et vice-versa. La littérature n'est cependant pas concluante sur ce point. Il sera intéressant d'examiner les facteurs qui influencent le nombre de revendications et de citations des brevets à la section 4.5 présentant les résultats de régressions.

## 4.4 Degré d'application

Un des aspects d'intérêt dans ce projet est l'effet de la durée de l'intervalle de temps entre le financement d'inventeurs-universitaires et le brevetage subséquent. Diverses méthodes de séparation sont utilisées pour comparer les intervalles courts et longs, i.e. en effectuant une ségrégation à partir de la moyenne, de la médiane et selon le quartile.

Tableau 4-7: Degré d'application de connaissances associées aux brevets en boucles selon l'intervalle de temps entre le financement et le dépôt de brevet

		NANOTECHNOLOGIE		BIOTECHNOLOGIE		NANOBIOTECHNOLOGIE	
		Moyenne (Dév. Std.)	n	Moyenne (Dév. Std.)	n	Moyenne (Dév. Std.)	n
Séparation par moyenne	Categorie						
	T < moyenne	2.92 (0.32)+	16	3.36 (0.75)+	47	3.42 (0.49)+	59
	T > moyenne	2.82 (0.41)+	24	3.38 (0.54)+	30	3.25 (0.58)+	20
	Total	2,87	40	3,37	77	3,38	79
Séparation par la médiane	Categorie						
	T < médiane	2.94 (0.27)+	13	3.33 (0.78)+	34	3.42 (0.48)+	54
	T > médiane	2.83 (0.42)+	27	3.40 (0.57)+	43	3.28 (0.57)+	25
	Total	2,87	40	3,37	77	3,38	79
Séparation par quartile	Categorie						
	4ième quartile	3.03 (0.25)+	5	3.35 (0.72)+	25	3.64 (0.37)**	29
	1ier quartile	2.89 (0.45)+	18	3.47 (0.54)+	24	3.12 (0.63)**	11
	Total	2,92	23	3,4	49	3,38	40

+ for  $p > 0.10$ , \* for  $p < 0.10$ , \*\* for  $p < 0.05$ , \*\*\* for  $p < 0.01$ , \*\*\*\* for  $p < 0.001$

Il est apparent qu'il n'y a pas de preuve pour supporter la prémisse selon laquelle un haut degré d'application des connaissances générées est associé à un plus petit intervalle de temps entre le financement et le brevetage, ce qui ne permet pas de supporter l'hypothèse H4a. En d'autres mots, les connaissances publiées par des chercheurs (autour du moment auquel ils reçoivent des fonds de firmes qui mènent à une invention brevetée basée sur ces connaissances) le sont dans des journaux qui portent le même niveau d'application, peu importe le sous-domaine en question. Il y a cependant une exception concernant les quartiles inférieurs et supérieurs de brevets nanobiotechnologiques brevetés. Les brevets octroyés plus récemment suivant l'obtention du financement (i.e. avec un intervalle plus court) sont associés à de la recherche plus théorique au moment du contrat. Ceci est contre intuitif, cependant puisque l'échantillon employé est relativement petit, il est plausible de supposer qu'il n'y a pas d'effet temporel, du moins pour l'instant.

## 4.5 Régressions

Cette section contient l'analyse des blocs de régressions. Elle contient les études suivantes :

- (1) Analyse du bloc A : le fait de breveter combiné au nombre de brevets du chercheur :
- (2) Comparaison entre blocs B et C : le nombre de citations reçues après 5 ans par rapport au nombre de revendications, chacun combiné au fait de breveter;
- (3) Analyse du bloc D : le nombre de brevets du chercheur les années où il brevète;
- (4) Comparaison entre blocs E et F : le nombre de citations reçues après 5 ans par rapport au nombre de revendications.

En conséquence, les régressions des blocs A à C se font sur un échantillon d'environ 8770 chercheurs-années, alors que les régressions des blocs D à F se font sur un échantillon beaucoup plus restreint correspondant seulement aux années où les chercheurs ont des brevets (soit environ 420 observations).

Avant de continuer, mentionnons que puisque certaines des variables sont transformées pour obtenir des modèles plus représentatifs de la réalité, il devient nécessaire d'interpréter leurs coefficients en conséquence. Les cas suivants sont observés :

- Variable dépendante non-transformée et variable explicative non-transformée : ayant l'équation suivante  $= \beta_0 + \beta_1(x) + \varepsilon$  , une croissance d'une unité de la variable explicative engendre un effet de  $\beta_1$  sur la variable dépendante.
- Variable dépendante transformée en logarithme naturel et variable explicative non-transformée : ayant l'équation suivante  $= \beta_0 + \beta_1(x) + \varepsilon$  , une croissance d'une unité de la variable explicative engendre une croissance de  $\exp(\beta_1)$  sur la variable dépendante.
- Variable dépendante non-transformée et variable explicative transformée en logarithme naturel : ayant l'équation suivante  $y = \beta_0 + \beta_1(\ln x) + \varepsilon$ , une croissance de 1% de la variable dépendante  $x$  engendrera une croissance de  $\beta_1 * (\ln 1,01)$ <sup>44</sup>
- Variable dépendante et explicative transformées en logarithme naturel : ayant l'équation suivante  $= \beta_0 + \beta_1(\ln x) + \varepsilon$  , une croissance de 1% de la variable explicative engendre une croissance de  $\beta_1 * (1,01)$ . Ceci s'explique par le fait que les logarithmes s'annulent, donc aucune transformation supplémentaire n'est requise (UCLA ATS, 2011).

#### 4.5.1 Analyse du bloc A - breveter et nombre de brevets

Nous allons parcourir ici les facteurs que nous soupçonnons d'influencer le fait qu'un chercheur universitaire participe au développement d'une invention brevetée et le nombre de brevets auxquels il collabore. Les tableaux utilisés pour faire ces inférences se trouvent à l'annexe 14.

Tel que mentionné précédemment, quatre types de régressions ont été estimés afin de s'assurer de bien identifier les facteurs qui influencent le fait de breveter ainsi que le nombre de brevets obtenus. Les deux premiers modèles (Tableaux 5-18 et 5-19) utilisent des régressions à binomiale négative, la seconde utilisant un modèle de panel (xtnbreg) et la première sans le panel mais en

---

<sup>44</sup>Expliqué par le fait qu'une croissance de 1% implique  $y(x + 1\% * x) - y(x) = \beta_1 [\ln(x + 1\% * x) - \ln(x)] = \beta_1 \left[ \ln \frac{x+1\%*x}{x} \right]$

tenant compte de la non-indépendance des observations pour un même chercheur au fil du temps (nous ajoutons la commande *vce cluster id* aux régressions utilisant la procédure *nbreg*).

Ces deux premiers modèles sont consistants quant à l'influence positive des contrats sur le fait de breveter et sur le nombre de brevets obtenus. Il en est de même pour l'effet de clique, qui a un effet non linéaire sur le fait de breveter. Il semble en effet que plus l'effet de clique est important, plus les chercheurs ont tendance à déposer des brevets, mais que cette influence positive a une limite. Au-delà de cette limite, l'effet positif de la clique s'effrite. Il serait donc souhaitable que les chercheurs puissent collaborer avec un certain nombre de communautés relativement distinctes, plutôt qu'avec une seule clique bien intégrée afin de breveter.

En revanche, il ne semble pas que le financement public ait une influence induite sur le fait de breveter puisque la variable *ltotsubvmoyan3* n'est pas systématiquement significative. Le montant des subventions devient significatif lorsque nous tenons compte de l'endogénéité pour cette variable dans les données de panel (*xtnbreg*). L'annexe 20 (Tableau 5-41) présente les régressions des premières équations sur les instruments qui expliquent le montant des subventions, variable endogène dans la seconde équation.

Évidemment, si un chercheur a déjà breveté dans le passé [*avg\_pat\_past*], la probabilité qu'il brevète à l'année *t* est plus élevée. Cela est d'autant plus vrai pour le fait d'avoir déjà reçu un financement qui a mené à un brevet [*nbloop*]. Il s'agit ici en quelque sorte d'un effet productif et collaboratif cumulatif.

Le nombre d'articles produits par les inventeurs universitaires n'ont vraisemblablement pas d'effet significatif sur le fait de breveter. Les effets linéaires et quadratiques ont été testés et ne révèlent rien de tangible que nous puissions utiliser pour effectuer des inférences à ce niveau.

Lorsque nous considérons le modèle de régression à binomiale négative à zéro-augmenté (*zinb* au Tableau 5-20), les résultats sont sensiblement les mêmes au niveau des brevets passés et de l'effet de clique, bien que dans ce dernier cas, le degré de signification soit moindre. Il est toutefois intéressant de noter que ce modèle nous permet de distinguer les chercheurs qui auraient la 'capacité' de breveter de ceux qui n'ont pas cette inclinaison. En effet, la partie augmentée (*inflate*) de la régression permet d'expliquer les facteurs qui contribuent au fait que les chercheurs ne brevètent pas. Moins un chercheur a de financement privé, moins il aura une propension à breveter. Dans cette partie de la régression, il peut sembler surprenant qu'un chercheur dont les

articles sont plus appliqués aura moins tendance à breveter. Cela peut sembler contre-intuitif. Les deux précédents modèles de régressions (nbreg et xtnbreg) montrent un effet positif de la variable sur le fait de breveter, ce qui veut dire que plus les articles sont de nature fondamentale, plus un chercheur a tendance à breveter. Une explication possible de ce phénomène serait un effet de substitution entre les publications à caractère plus appliqué et les brevets, qui sont par définition des connaissances plus appliquées. Soit les chercheurs choisissent délibérément de mettre ces connaissances plus appliquées dans le domaine public (articles) afin d'en faire bénéficier la société, ou ils (ou le bureau de la recherche de l'université) n'entrevoient pas de débouchés industriels immédiats, donc ils ne brevètent pas, soit ils décident de breveter. Les deux choix semblent mutuellement exclusifs. L'arbitrage brevet-article ne se ferait donc pas au niveau des connaissances plus fondamentales, mais au niveau des connaissances et des technologies appliquées.

Les variables témoins des années illustrent une réduction significative du brevetage des universitaires vers la fin des années étudiées (2003-2005) et aussi une réduction entre 1997 et 1999. Ceci vient supporter la forme de cloche que possèdent les brevets universitaires illustrée dans la section de résultats précédente. Il faut aussi mentionner que le délai observé entre la date de dépôt et d'octroi des brevets s'est considérablement allongé récemment, ce qui contribue à la baisse apparente de 2003-2005. Il se peut qu'une partie des brevets n'aient pas encore été octroyés et n'apparaissent donc pas dans notre échantillon.

#### **4.5.2 Comparaison entre les blocs B et C - breveter avec revendications et citations**

Lorsque l'on remplace la variable dépendante des régressions précédentes [nbpat] par le nombre de revendications (bloc B, annexe 15) et de citations (bloc C, annexe 16) et que l'on utilise les mêmes modèles, certains résultats demeurent (principalement pour les modèles nbreg et xtnbreg), mais d'autres apparaissent (pour les modèles zinb). Par exemple, le degré d'application des articles des chercheurs demeure positif et significatif et ce tant pour les revendications que pour les citations. Ceci implique qu'un brevet dont les connaissances associées sont à caractère fondamentale a tendance à recevoir plus de citations et avoir plus de revendications. L'importance de la proximité à la science fondamentale dans les domaines de nanotechnologie et de biotechnologie est donc en quelque sorte confirmée ici. Les contrats sont aussi toujours un

facteur d'influence des brevets que ce soit au niveau des citations ou des revendications. Il en est de même pour le fait d'avoir breveté dans le passé. Même si le fait de publier des articles dans les années menant au dépôt de brevet n'avait pas d'influence sur la qualité seule des brevets, elle en a une positive sur le fait de breveter et que ce brevet ait de nombreuses revendications et citations (surtout avec le xtnbreg). Ceci vient confirmer en quelque sorte un corollaire des propos de Gulbrandsen & Smeby (2002) stipulant l'importance des liens entre industrie et université pour la production scientifique. De manière plus générale, les catégories semblent avoir une influence sur les variables dépendantes. Le fait d'œuvrer (incluant publications etc.) en nanotechnologie influence de manière négative le fait de breveter et avoir de la qualité par rapport à la nanobiotechnologie.<sup>45</sup> Le fait d'œuvrer en biotechnologie a un aussi un effet significatif négatif sur la propension de breveter avec beaucoup de citations ou revendications. Encore une fois, ceci est par rapport à la nanobiotechnologie (la variable omise). Ceci pourrait être dû à la multidisciplinarité pointue de la nanobiotechnologie, ce domaine étant à l'intersection de la nanotechnologie et de la biotechnologie. Tout comme pour les régressions précédentes, les variables témoins des années récentes (2003-2005) ont un effet significatif négatif par rapport à l'année de référence (1996). Ceci est cependant vrai pour les deux variables dépendantes examinées. Plus un brevet est récent, moins il reçoit des citations et moins il possède de revendications.

La densité égocentrique d'un chercheur académique a généralement un effet significatif tant sur les citations que sur les revendications. Ainsi un chercheur dont le réseau académique est bien connecté a plus tendance à déposer des brevets de qualité. L'effet est cependant quadratique, ainsi si le réseau du chercheur est trop inter-relié ou trop « fermé » cela affecte négativement la qualité de ses brevets.

Une des seules différences entre les coefficients estimant les citations et les revendications se trouve au niveau des subventions pour les régressions à binomiale négative (avec et sans panel). Les subventions d'un chercheur n'ont pas d'influence sur sa capacité à émettre un brevet qui sera

---

<sup>45</sup> Nous avons essayé à plusieurs reprises d'inclure les sous-catégories de la nanotechnologie et de la biotechnologie dans les Blocs C, D, E et F, mais cela produit systématiquement des cas de non-convergence du maximum vraisemblance. Ainsi, une étude plus approfondie est requise à ce niveau.

cit  , toutefois il y a un effet g  n  ralement positif sur le fait de brev  ter avec plusieurs revendications. Un effet quadratique est d  tect  , mais seulement avec la r  gression    binomiale n  gative.

Les autres diff  rences principales entre les r  sultats des citations et des revendications apparaissent surtout dans les mod  les de r  gression    binomiale n  gative    z  ro-augment   (zinb). Tout d'abord, lorsque nous comparons la partie de la r  gression qui s  pare les inventeurs potentiels des autres non inventeurs (*inflate*), les contrats jouent un r  le beaucoup plus important dans le cas des revendications que dans le cas des citations. En outre, l'effet du niveau plus fondamental de la recherche a aussi un effet positif sur le fait de brev  ter des inventions avec un plus grand nombre de revendications. Alors que le fait d'avoir brev  t   dans le pass   a une forte influence sur la propension    g  n  rer des brev  ts avec un plus grand nombre de revendications, ce n'est pas le cas pour le nombre de citations. Dans le m  me mod  le, si nous consid  rons maintenant la partie principale de la r  gression, le fait d'avoir un plus grand nombre de brev  ts provenant de boucles d'innovation a un effet positif sur le nombre de revendications obtenues alors que cette variable n'a aucun effet sur le nombre de citations obtenues par ces brev  ts. Deux explications sont possibles    cet   gard. Il se peut que les brev  ts des chercheurs fassent aussi l'objet de liens-universit  -entreprise au sein de boucles d'innovation (nous pourr  ons examiner cet effet dans les r  gressions des blocs E et F un peu plus loin). Il se peut aussi que les chercheurs apprennent au contact des entreprises de par leurs relations avec elles par le pass   et que cons  quemment, les brev  ts qu'ils produisent aient une plus grande port  e (un plus grand nombre de revendications).

### 4.5.3 Analyse du bloc D – nombre de brev  ts

La mod  lisation du nombre de brev  ts auxquels participe un chercheur s'est effectu  e avec la r  gression    binomiale n  gative et la tobit    l'aide d'un   chantillon plus restreint, ne consid  rant que les ann  es o   les chercheurs ont des brev  ts. Les r  sultats de ces deux mod  les sugg  rent des r  sultats int  ressants concernant le financement. Les contrats semblent avoir un effet soit insignifiant ou n  gatif (lorsque l'endog  n  it   est mod  lis  e sur les articles [nbarticlemoyan3] ou les contrats [ltotcontmoyan3] dans la r  gression    binomiale n  gative) sur la propension d'un chercheur    avoir plus d'un brevet par ann  e; ce qui est relativement surprenant. Pour ce qui est du financement public, les subventions ont un effet lin  aire simple sur le nombre de brev  ts d'un



chercheur. Ainsi, un chercheur qui ne reçoit pas ou peu de subventions a davantage tendance à avoir plusieurs brevets et ceux qui en reçoivent beaucoup ont aussi tendance à avoir plusieurs brevets.

Pour ce qui est de son réseau académique, le chercheur dont la clique est trop bien intégrée et qui n'est pas assez ouvert à collaborer avec différents groupes de scientifiques réduit sa capacité à produire un plus grand nombre de brevets. En effet, l'effet de clique (*cliquishness*) du chercheur a une influence quadratique inverse, en forme de U-inversé, ce qui implique qu'une trop grande interconnexion du réseau académique de ce dernier entraîne alors une diminution du nombre de brevets auxquels il participe (ou contribue).

Le fait d'avoir participé à des boucles d'innovations dans le passé a un effet positif et significatif pour le nombre de brevets. Plus un chercheur a une expérience de collaboration fructueuse avec l'industrie, plus il sera en mesure de produire de nombreux brevets. De manière cyclique, son nombre de brevets va inciter les firmes ayant déjà collaboré avec lui à refinancer et à collaborer de nouveau avec ce chercheur. C'est un cycle important qui pourrait principalement tirer ses origines de l'effet Saint-Matthieu. Le degré d'application des articles publiés n'a encore une fois aucun effet sur le nombre de brevets que le chercheur produit. Les années ont des effets généralement peu ou pas significatifs. Une exception est toutefois notable pour les années 1997-2001 qui ont un effet négatif et significatif par rapport à 1996 dans la régression tobit. Il semble que le fait d'être à l'Université Laval ou de Sherbrooke ait un impact négatif sur le nombre de brevets, par rapport à McGill. Le fait d'être à l'Université Concordia a un effet positif. Concordia a une grande spécialisation en NEMS, ce qui pourrait expliquer ce fait.

#### **4.5.4 Comparaison entre les blocs E et F – nombre de revendications et de citations**

L'analyse des blocs E (annexe 18) et F (annexe 19) permet de cibler spécifiquement les éléments causant une divergence des mesures de qualité, soit le nombre de revendications et de citations. Il faut mentionner que la comparaison des coefficients et leur significativité est possible entre la régression par moindres carrés ordinaires sur [nbclaims] et celle de tobit sur [nbcitp5] car les logarithmes naturels sont calculés pour chacun. Un des seuls éléments ayant un effet exclusivement positif et significatif tant sur le nombre de revendications que sur le nombre de

citations est le fait d'être d'avoir bénéficié d'une boucle d'innovation. En effet, le fait que l'inventeur ait reçu des fonds de l'organisation qui possède les droits sur l'invention a un effet positif et significatif tant sur le nombre de revendications que sur le nombre de citations. La liaison industrie-université est telle nous l'avions soupçonné, c'est-à-dire importante pour la qualité des inventions brevetées.

Plusieurs éléments expliquent cependant une divergence entre les facteurs qui influencent les deux variables dépendantes à l'étude : le financement et les effets de réseau. Premièrement, les contrats expliquent le nombre de revendications pour les chercheurs ayant au moins un brevet. En effet, nos résultats suggèrent qu'il y a un effet quadratique avec  $\beta_{L\_ltotcontmoyan3} = -0.0661$  et  $\beta_{L\_sq\_ltotcontmoyan3} = 0.0068$ <sup>46</sup> croissant lorsque la moyenne des contrats sur les 3 années précédant l'application du brevet dépasse \$129 CAD. Il semblerait donc qu'un chercheur recevant des contrats supérieurs à cette somme, ait une propension à développer un brevet d'une plus grande étendue. Cette somme étant très petite nous pouvons l'interpréter comme un effet exponentiel positif, i.e. le fait d'avoir des montants de contrats plus élevés mène à davantage de revendications. En revanche, les contrats n'ont pas d'effet sur le nombre de citations. Pour leur part, les subventions reçues par un chercheur ne peuvent expliquer la propension de ce dernier à avoir un grand nombre de revendications. Les subventions sont cependant utiles dans la modélisation sans endogénéité pour expliquer le nombre de citations, l'effet est linéaire<sup>47</sup> et négatif, donc plus un chercheur reçoit du financement public, moins ses brevets reçoivent de citations.

Deuxièmement, nous abordons les effets réseaux. La densité égocentrique (*cliquishness*) joue un rôle important pour déterminer si un chercheur brevète avec revendications importantes, mais un peu moins pour les citations. Par ailleurs, il y a un effet quadratique non-négligeable lorsque l'on considère l'endogénéité sur les subventions, ce qui suggère qu'un effet de clique académique trop important engendrerait une diminution du fait d'avoir de nombreuses revendications. Le nombre

---

<sup>46</sup> Exemple avec la binomiale négative (nbreg), mais les autres coefficients sont du même ordre de grandeur, donc l'exemple est représentatif.

<sup>47</sup> Nous avons testé des effets quadratiques pour le montant des subventions sans succès. L'effet est donc considéré linéaire.

de revendications associées à un brevet est directement affecté par la proportion d'inventeurs universitaires ayant contribué à l'invention et mentionnés sur le brevet comme tel. Cette proportion a un effet négatif et significatif, c'est-à-dire que plus il y a d'universitaires dans une équipe, moins il y a de revendications au brevet produit. Ceci est justifiable par le fait que les revendications sont en général d'ordre plus appliqué ou industriel. Encore une fois, il n'y a pas d'effet significatif de cette variable sur le nombre de citations.

Certains facteurs n'ont tout simplement pas d'influence sur le nombre de citations ou de revendications. C'est le cas du nombre d'articles publiés par un chercheur dans les années précédant le dépôt de brevet (qui n'est pas significatif dans nos modèles). Ainsi, le seul fait de publier n'a pas d'influence sur le nombre de citations ou sur le nombre de revendications des brevets d'un chercheur universitaire. Il en va de même pour le degré d'application qui n'a pas d'effet significatif sur la quantité de citations ou de revendications. Ceci pourrait impliquer que la nature de la recherche effectuée en nanotechnologie et en biotechnologie par un chercheur n'influence pas la qualité de ses brevets. Ceci vient donc renforcer des résultats trouvés plus haut concernant le degré d'application.

Concernant les sous-domaines des brevets, les résultats se trouvent dans les tableaux 5-35 et 5-36 de l'annexes 18 et les tableaux 5-39 et 5-40 de l'annexe 19. Le fait d'avoir un brevet en nanotechnologie a un effet positif significatif (par rapport au groupe de référence, i.e. la nanobiotechnologie) tant sur le nombre de revendications que sur les citations. Cet effet est principalement attribuable aux catégories de nanodispositifs et nano-outils et de nanostructures et nanomatériaux. Le résultat obtenu ici est comparable à ce qui était trouvé par Hullman (2007) portant sur la spécialisation régionale dans les nanodispositifs et les nanomatériaux. Le fait d'être en biotechnologie a un effet négatif sur le nombre de citations que reçoivent les brevets. Il n'y a pas d'effet sur le nombre de revendications cependant. En analysant de plus près les sous-catégories, nous remarquons que la seule n'ayant pas d'effet différent par rapport à la nanobiotechnologie est celle d'ADN /ARN. Les autres, i.e. protéines et autres molécules, culture et ingénierie cellulaire et tissulaire, les techniques de processus biotechnologiques, les vecteurs de gènes et d'ARN et la bioinformatique ont un effet significatif par rapport à la nanobiotechnologie.

## CHAPITRE 5 DISCUSSION GÉNÉRALE

Plus tôt dans cet article nous avons décidé de vérifier cinq hypothèses. Les quatre premières concernent les différences et les similitudes entre la biotechnologie et la nanotechnologie pour commencer et ensuite entre les brevets qui sont le résultat de l'implication des entreprises dans la recherche universitaire, via le financement direct de cette recherche, et les autres brevets dans ces domaines. La dernière porte sur nos attentes des résultats de régression. Examinons ces sept hypothèses, à leur tour.

La première hypothèse propose que la nanotechnologie et la biotechnologie ne suivent pas un schéma de développement similaire malgré le nombre non-négligeable de similarités entre ces domaines<sup>48</sup>. Nous trouvons toutefois que la nanotechnologie est différente à certains égards de la biotechnologie. Les brevets en nanotechnologie ont généralement plus de revendications et plus de citations après 5 ans que ceux en biotechnologie (excluant le sous-domaine de la nanobiotechnologie). Par ailleurs, la recherche effectuée par les inventeurs universitaires, dans les 3 années précédant la demande de brevet est plus appliquée en nanotechnologie qu'en biotechnologie. Pour ces raisons, l'hypothèse H1 est supportée.

Une des raisons de comparer la biotechnologie et la nanotechnologie est l'émergence du domaine de la nanobiotechnologie. Nous avons postulé que la nanobiotechnologie suit de manière plus proche la biotechnologie que la nanotechnologie. Les données examinées sont en mesure de rejeter cette hypothèse, quoiqu'à certains égards, la nanobiotechnologie est différente des deux domaines. Par exemple, la moyenne de temps entre le financement et l'application des brevets et ensuite l'octroi des brevets est beaucoup plus petite pour la nanobiotechnologie et les montants de financement par contrat ou subvention sont relativement plus grands. En général, la nanobiotechnologie est différente de la nanotechnologie en termes de revendications, de citations et du degré d'application de la recherche à 3 années de l'application du brevet. La comparaison avec la biotechnologie donne plusieurs résultats : pour le nombre de revendications et le degré d'application de la recherche, la nanobiotechnologie est différente de la biotechnologie. En

---

<sup>48</sup> Par exemple, la proportion d'inventeurs académiques impliqués en brevetage suit un trajet similaire à quelques années d'intervalle.

termes du nombre de citations, la différence n'est pas statistiquement significative. En général cependant, l'hypothèse H2 est rejetée.

Les deux dernières hypothèses portent sur les boucles d'innovation relatives au financement/brevetage. Le nombre de ces événements (et sa proportion par rapport au nombre total de brevets) est étonnamment négligeable, avec très peu de ces organisations étant autre chose que des firmes. Il y a plusieurs raisons pour expliquer cela. Premièrement, les organisations qui désirent développer des innovations ne s'intéressent pas autant que l'on puisse croire à la recherche universitaire. Deuxièmement, les scientifiques reçoivent des fonds de diverses organisations qui n'espèrent pas avoir un brevet comme extrant, e.g. les sociétés financent des scientifiques afin de maintenir des liens avec le milieu universitaire et, plus important avec la science fondamentale tel que mentionné par Meyer (2006) et les universités pourvoient des fonds pour leurs scientifiques pour les aider à accroître leurs publications. Troisièmement, ce n'est pas tant la collaboration financière qui importe pour le brevetage, mais plutôt la collaboration de connaissances, même s'il est difficile d'imaginer des universitaires dans les domaines de la nanotechnologie et de la biotechnologie innover sans financement approprié.

La troisième hypothèse suggère que les brevets générés par ces boucles sont de plus grande qualité que ceux générés hors boucle, la qualité étant mesurée par le nombre de revendications et le nombre de citations obtenus par ces brevets. Les résultats pour la biotechnologie suggèrent qu'il y a une différence significative entre les brevets dans et hors boucle : les brevets en boucles ont un plus grand nombre de revendications, mais un plus petit nombre de citations. Ceci a été à l'origine de la question concernant le meilleur indicateur de qualité. Une question qui n'a pas encore tout à fait été résolue selon notre compréhension de la littérature. Pour le domaine de la nanobiotechnologie, il n'existe pas de différence significative. Il est important de mentionner par contre que globalement les résultats de nos régressions suggèrent que le fait d'avoir participé à une boucle d'innovation ou d'avoir déjà participé à une boucle d'innovation dans le passé augmentent le taux de citations et de revendications. Ainsi, l'hypothèse H3 n'est que partiellement supportée.

Notre analyse intra boucle et extra boucle supporte l'hypothèse H1 aussi, c'est-à-dire qu'il y a une différence significative entre la nanotechnologie et la biotechnologie et l'hypothèse H2 (impliquant un rapprochement entre la nanobiotechnologie et la biotechnologie). Même si la

nanotechnologie et la biotechnologie partagent des points communs multiples, voyant comment la taille et la nature des boucles diffèrent en fonction de leurs domaines respectifs, ce pourrait être une erreur de reproduire le processus de connaissance de la biotechnologie dans les nanotechnologies car elle ne semble pas être aussi efficace. En fait, le processus de création de connaissances en nanotechnologie posséderait son propre modèle d'innovation optimal. Ainsi, afin d'extraire des réponses constructives concernant cette question, des recherches supplémentaires sont nécessaires.

La quatrième hypothèse n'a pas révélé des résultats intéressants en soi. Nous avons étudié le lien entre le décalage dans le temps entre le financement et le brevetage et le degré d'application de la recherche effectuée au moment du financement. Nous trouvons que l'hypothèse H4a est supportée. Par contre, nous trouvons que la nanotechnologie est généralement plus appliquée que la biotechnologie.

Pour ce qui est de notre objectif secondaire, l'hypothèse H5 se trouve à être partiellement validée. Avec la comparaison des indicateurs de citations et de revendications nous trouvons que les éléments influençant la quantité de chacun sont le fait de participer à une boucle d'innovation, le fait d'être en nanotechnologie et les années d'applications. Individuellement, le nombre de revendications est influencé par les contrats reçus, la densité égocentrique des chercheurs, la proportion d'inventeurs universitaires. Pour leur part les citations sont plutôt corrélées avec les subventions reçues et le fait d'être en biotechnologie.

De plus, nous remarquons qu'il y a des différences significatives entre les coefficients des modèles E&B et C&D. En effet, il semble que certains des éléments qui sont utiles pour caractériser la qualité d'un brevet ne soient pas les mêmes que ceux influençant le fait d'avoir un brevet et qu'il soit de qualité. Pour ce cas, les facteurs ayant un effet similaire sur les deux caractéristiques de qualité sont les contrats reçus, le degré d'application, l'effet de clique (*cliquishness*) du chercheur, les boucles d'innovation passées, et le domaine des brevets (nanotech ou biotech). Un élément qui influence positivement uniquement le fait de demander et d'obtenir un plus grand nombre de revendications sont les subventions.

## CONCLUSION ET RECOMMANDATIONS

### 5.1 Synthèse

L'objectif premier de cette recherche était de modéliser le processus de financement-brevetage des chercheurs universitaires québécois et d'examiner s'il y avait une différence notable entre celui observé en nanotechnologie et celui de la biotechnologie. Pour ce faire nous avons créé le concept de boucle d'innovations et avons analysé statistiquement plusieurs facteurs propres au modèle d'innovation. Nous avons ensuite créé des modèles pour illustrer l'impact des propriétés de réseaux, des chercheurs et des brevets passés pour évaluer la propension d'un universitaire à breveter. En somme, malgré leurs similarités (dues entre autres au fait qu'ils sont des domaines à bases scientifiques mais aussi par le fait qu'ils sont des percées grilichéennes) les deux technologies à l'étude ici sont des mondes à part. Même si les implications d'inventeurs universitaires dans le processus de brevetage indiquent des niveaux similaires tant en nanotech qu'en biotech, c'est seulement dans ce dernier cas qu'il y a une supériorité des revendications (mais moins de citations). Une note à mentionner est le fait que plusieurs des régressions ne supportaient pas les sous-catégorisations, rendant une analyse pointue de ces sous-domaines impossible. D'autres modèles économétriques compilés portaient sur la comparaison des mesures de qualités. Elles dressent le portrait de ce qui importe pour définir la qualité d'un brevet.

Il existe un certain nombre de limites à notre recherche, que nous décrivons ci-dessous.

### 5.2 Limitations

(1) Nous traitons l'intersection entre la nanotechnologie et la biotechnologie de manière très générale sans faire la distinction entre la nanobiotechnologie et la bionanotechnologie. Pour les fins de ce projet, ce n'était pas nécessaire. Par contre, il serait très intéressant de procéder à cette analyse en comparant les différentes définitions utilisées pour décrire chacun des domaines en question.

(2) La complexité des processus d'innovation en nanotechnologie comme en biotechnologie et leur non-linéarité rendent la réalisation d'une analyse complète laborieuse. C'est en partie pour cette raison que nous nous limitons à analyser les boucles d'innovation de chercheurs

universitaires. Or, même si cette méthode nous permet de faire des découvertes intéressantes et inattendues, elle offre une vision limitée en soi.

(4) Même si nous avons fait de grands efforts pour nous assurer que nos scientifiques ont été convenablement liés les uns avec les autres, il est fort probable que nous avons (dans une certaine mesure) inclus de faux positifs dans notre processus de bouclage et exclus de faux négatifs. Ceci est encore plus pertinent pour les cessionnaires pour lesquels il n'est pas rare d'être représentés au fil du temps par des noms différents. En effet, le nom d'un cessionnaire est utilisé pour associer ce dernier à lui-même dans la base de données de financement, cependant nous ne pouvons tenir compte du fait que même si les noms d'inventeurs restent permanents, le cessionnaire peut changer. Ainsi il est possible que nous ayons faussement identifié un évènement comme étant une boucle alors que ce n'était pas le cas.

(5) L'utilisation de données sur le co-brevetage et la co-publication devrait être abordée avec prudence car les « règles » pour déterminer les auteurs et les inventeurs peuvent varier selon le domaine (Ducor, 2000). Ceci n'est pas un problème pour l'instant car les domaines à l'étude étant de la haute technologie, les coutumes d'identification d'auteur sont similaires. Cependant, si nous décidons d'effectuer une analyse plus approfondie avec d'autres secteurs d'activité dans le futur, nous devrions en tenir compte.

(6) Les brevets sont utilisés comme mesure d'innovation or même si la nanotechnologie et la biotechnologie sont des domaines où le brevetage est le meilleur moyen de protection, il n'est pas le seul. Nous ne tenons pas compte des innovations qui sont tenues secrètes par exemple, même si elles sont rares<sup>49</sup>.

(7) La classification des brevets est différente en nanotechnologie et biotechnologie. Une option avait été suggérée pour effectuer une validation empirique des résultats de : ré-effectuer la classification de la nanotechnologie en utilisant la même méthode que la biotechnologie. Cependant les données sources n'existent pas sous le même format. Plus d'attention est requise à ce niveau pour des études subséquentes.

---

<sup>49</sup> Les contrats peuvent mener à de telles découvertes.



(8) La compilation des degrés d'application n'est pas exhaustive. En effet, il se peut qu'un chercheur ait publié dans une revue qui n'a pas été recensée et classée par *CHI Research*. Ainsi, il y a un biais possible à considérer à ce niveau.

### 5.3 Implications sur la politique

Avec ce projet nous avons cartographié une portion du processus d'innovation impliquant les chercheurs universitaires et il est pertinent de mentionner que ce travail peut être consulté à titre informatif dans l'application de la politique industrielle mais aussi pour ceux développant les politiques institutionnelles internes de R&D. En particulier, un point important à soulever est la pertinence de l'association d'une firme en nanotechnologie comme en biotechnologie avec le milieu universitaire pour le brevetage. Quoique ne présentant que des résultats immédiats limités<sup>50</sup>, cette association est très utile pour prédire la performance de brevetage future des firmes<sup>51</sup>. Ceci n'est donc pas à négliger lors de l'élaboration des stratégies industrielles de R&D.

Les politiques nationales de développement du secteur public et des institutions devraient également en tenir compte afin d'envisager combien et à qui ils doivent accorder des fonds de recherche.

### 5.4 Directions de recherche

La nanotechnologie, à cause de sa nature hautement tacite (Schiffauerova & Beaudry, 2009), est un domaine particulier. En effet, ce secteur survit en grappes et en réseaux en partie parce que les connaissances nécessaires pour y faire des avancements technologiques considérables sont tacites, c'est-à-dire non-formalisées. Ainsi, le degré d'innovation en nanotechnologie n'est pas simplement une conséquence du nombre de dollars qu'on y injecte, mais aussi, et en grande partie, du dialogue entre les ingénieurs, chercheurs et professionnels du

---

<sup>50</sup> En termes du nombre de brevets (cf. section 4.2)

<sup>51</sup> Cf. section 4.5

domaine. L'interaction qu'ont ces différents acteurs est d'une importance primordiale, car sans elle, il n'y aurait pas de nanotechnologie au Canada ou du moins, elle ne serait pas ce qu'elle est aujourd'hui. Ainsi, un de nos objectifs futurs sera de modéliser le réseau de brevetage des chercheurs universitaires et de l'intégrer à notre étude en cours.

Nous tenons également à déterminer si oui ou non les étudiants universitaires qui ont reçu du financement des entreprises industrielles ont continué à travailler pour eux après leurs études supérieures et par la suite développé une invention brevetée. Cela donnerait une bonne indication quant à l'importance à long terme du financement dans le développement de réseaux scientifiques en biotechnologie et en nanotechnologie. En bref, il y a de nombreuses possibilités pour de futures investigations et d'autres recherches sont évidemment nécessaires afin de bien saisir la relation entre le financement et l'innovation dans les domaines de haute technologie que sont les nanotechnologies et les biotechnologies.

## **BIBLIOGRAPHIE**

Holdridge, G.M. (2006). The role of the U.S. Government in Nanoscale Science and Technology. In Foster L.E., Nanotechnology: Science, Innovation and Opportunity (pp.63-76). Upper Saddle River, N.J.: Prentice Hall.

David, K. & Thompson, P.B. (2008). What can nanotechnology learn from biotechnology (2008). Burlington, MA: Elsevier.

## **RÉFÉRENCES**

Aizawa, A. (2003). “An information theoretic perspective of tf-idf measures”. Information processing and management. Vol 39-1. 45-65.

Arundel, A. (1999) Diffusion of biotechnologies in Canada: Results from the survey of biotechnology use in Canadian Industries. Statistics Canada.

Banerjee, P., Gupta, B. M. & Garg, K. C. (2000). Patent statistics as indicators of competition, an analysis of patenting in biotechnology. Scientometrics. Vol 47. No1. 95-116.

Bardhan, Ashok and Kroll, Cynthia A. (2003) The New Wave of Outsourcing. Fisher Center for Real Estate & Urban Economics Research Report Series No. 1103. Available at SSRN: <http://ssrn.com/abstract=985741>

Barirani, A. (2008) Mémoire de Maîtrise intitulé « PatentBot: Un extracteur de données sur le web ». École Polytechnique de Montréal.

Beaudry, C. & Clerk-Lamalice, M. (2009). Impact of Quebec University Research Financing in Biotechnology, 49ième Congrès Annuel de la Société Canadienne de Science Économique. 1-11.

Beaudry, C. & Clerk-Lamalice, M. (2010). Grants, contracts and networks: What influences biotechnology scientific production, Danish Research Unit for Industrial Dynamics (DRUID). Conference, London, 16-18 June, 31 pages.

Bell, G. G. (2005). Clusters, Networks and Firm Innovativeness, Strategic Management Journal 26, 287-295.

- Bidanda, B., Arisoy, O., Shuman, L.J. (2006). Offshoring manufacturing: Implications for engineering jobs and education: A survey and case study. *Robotics and Computer-Integrated Manufacturing*. Volume 22, Issues 5-6, Pages 576-587.
- Blackman, M. (1995) Provision of patent information: A National Patent Office Perspective, in *World Patent Information*, Vol. 17(2). 115-123
- Bonaccorsi, A. & Piccaluga, A. (1994). A theoretical framework for the evaluation of university-industry relationships, *R&D Management* 24(3), 229-247.
- Bonaccorsi, A., & Thoma, G. (2007). Institutional complementarity and inventive performance in nano science and technology. *Research Policy* 36, 813–831.
- Bozeman, B., Laredo, P. & Magematin, V. (2007). Understanding the emergence and deployment of « Nano » S&T, *Research Policy* 36(6), 807-812.
- Breschi, S. & Catalini, C. (2010). Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks, *Research Policy* 39, 14-26.
- Breschi, S., Lissoni, F. & Montobbio, F. (2005). From publishing to patenting: do productive scientists turn into academic inventors ? *Revue Économique Industrielle*, No 110, 2ième trimestre, 1-28.
- Bresnahan, T., Gambardella, A. & Saxenian, A.L. (2001). Old Economy' Inputs for 'New Economy' Outcomes: Cluster Formation in the New Silicon Valleys Industrial and Corporate Change 10(4), 835-860.
- Burnham, K. P., & Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2ième édition. Springer-Verlag.
- Butler, J. (2011) Presentation theme: "Development of University-Based Entrepreneurship Ecosystems. Global Perspectives on Technology Transfer and commercialization in Building Innovative Ecosystems". Triple Helix IX International Conference, July 12th 2011, Stanford University.
- Calero, C., Buter, R., Valdés, C.C. & Noyons, E. (2006). How to identify research groups using publication analysis: an example in the field of nanotechnology, *Scientometrics* 66(2), 365-376.

- Casper, S. (2007). How do technology clusters emerge and become sustainable? Social network formation and inter-firm mobility within the San Diego biotechnology cluster, *Research Policy* 36, 438-455.
- Daim, T., Monalisa, M., Dash, P. & Brown, N. (2007). Time lag assessment between research funding and output in emerging technologies. *Foresight* 9(4), 33-44.
- Dalpé, R. (2002). Bibliometric Analysis of Biotechnology. *Scientometrics*. Vol 55, No 2, 189-213.
- Darby, M.R., & Zucker, L.G. (2003). Grilichesian breakthroughs: inventions of methods of inventing and firm entry in nanotechnology. NBER working paper series. National bureau of economic research. 1-62.
- De Nooy, W., Mrvar, A. & Batagelj, V. (2005). *Exploratory Social Network Analysis with Pajek*. Cambridge University Press.
- Diodato, V. (1994). *Dictionary of bibliometrics*. New York: The Haworth Press.
- Domeniconi, C., Barbara, D., Chaudhary, H., Al-Timimi, A. & Jamison, C.(2005). Data-Mining Techniques for Microarray data analysis. In M.M. Kantardzic & J. Zurada, *Next Generation of Data-Mining Applications*.(pp.295-330). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Dong, G., Li, J. & Wong, L. (2005). The use of emerging patterns in the analysis of gee expression prfiles for the diagnosis and understanding of diseases. In M.M. Kantardzic & J. Zurada, *Next Generation of Data-Mining Applications*.(pp.331-354). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Fall, C.J., Tôrcsvari, A., Benzineb, K., Karetka, G. (2003). “Automated categorization in the international patent classification”. *Special Interest Group on Information Retrieval*. Vol 37-1. 10-25.
- Fishbine, G. (2002). *The Investor’s Guide to Nanotechnology & Micromachines (First Edition)*. New York: John Wiley & Sons, inc.
- Franks, A. (1987). Nanotechnology, *Journal of Physics E-Scientific Instrumentation*. 1442-1451.
- Gagné, R. (2010) Note de cours “Économétrie Appliquée”. École des Hautes Études Commerciales de Montréal.

- Gans, J. S., Murray, F. E. & Stern, S. (2008). Patents, Papers, Pairs & Secrets: Contracting over the disclosure of scientific knowledge, Preliminary paper. 1-51.
- Gazit, E. (2007). Plenty of Room for Biology at the Bottom: An Introduction to Bionanotechnology (1er edition). Imperial College Press. 200pp.
- George, G., Zahra, S.A., Robley Wood, D. (2002). The effect of business-university alliances on innovative output and financial performance: a study of publicly traded biotechnology firms, *Journal of business Venturing* 17, 577-609.
- Geuna, A. (1997). Allocation of funds and research output: the case of UK Universities. *Revue d'économie industrielle*, Vol. 79. 1ier trimestre, 143-162.
- Gingras, Y., Godin, B. & Trépanier, M. (1999). La place des universités dans les politiques scientifiques et technologiques canadiennes et québécoises L'état québécois et les universités – Acteurs et enjeux. Presses de l'Université du Québec, pp. 69-100.
- Godin, B. et Gingras, Y. (2000). The place of universities in the system of knowledge production, *Research Policy* 29, 273-278.
- Godin, B. et Trépanier, M. (1995). La politique scientifique et technologique québécoise : la mise en place d'un nouveau système national d'innovation. *Recherches sociographiques*, vol. 36, n° 3, 1995, p. 445-477.
- Grossman, D. A. & Frieder, O. (1998) *Information Retrieval: Algorithms and Heuristics* (The Springer International Series in Engineering and Computer Science). 1ière Édition. Dordrecht, Pays-Bas: Springer.
- Gulbrandsen, M. & Smeby, J.-C. (2005). Industry Funding and university professors' research performance *Research Policy* 34, 932-950.
- <http://onlinelibrary.wiley.com/doi/10.1057/emr.2008.9/pdf>
- Hall, B.H., Jaffe, A., Trajtenberg, M. (2001). "Market Value and Patent Citations: A first look." UC Berkeley Working Paper No. E01-304.
- Hullman, A. (2006). The economic development of nanotechnology – An indicators based analysis European Commission, DG Research. 1-34.

- Hullman, A. (2007). Measuring and assessing the development of nanotechnology. *Scientometrics*, 70(3), 739-758.
- Igami, M. (2008). Exploration of nanotechnology via mapping of patent applications. *Scientometrics*, 77(2), 289-308.
- Jaffe, A.B., Trajtenberg, M. & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, 108(3), 577-598.
- Kostoff, R.N., Koytcheff, R.G., & Clifford, G.Y.L. (2007). Global nanotechnology research metrics, *Scientometrics*, 70(3), 565-601.
- Kotsiantis, S.B. (2007) "Supervised Machine Learning: a review of classification techniques" In *Emerging artificial intelligence applications in computer engineering*. Maglogiannis, IOS Press.
- Lafontaine, C. (2006) Le Québec Nanotech : les discours publics en matière de nanotechnologie entre promotion et fascination. *Quaderni*, Volume 61, Numéro 61. pp. 39-53.
- Langlois, (1992) R.N. Langlois, Transaction costs economics in real time. *Industrial and Corporate Change*, 1 , pp. 99–127.
- Lanjouw, J. O., & Schankerman, M. (2004). Patent quality and research productivity: Measuring innovation with multiple indicators, *The Economic Journal*, 114, 441–465.
- Larivière, V. (2009) Notes de cours SHM1310. Atelier de méthodologie en scientométrie. Université du Québec à Montréal.
- Lissoni, F. & Montobbio, F. (2008). Inventorship and authorship in patent –publicatins pairs: an inquiry into the economics of scientific credit. *CESPRI*. pp. 1-45.
- MacNaghten, P. (2008). From Bio to Nano: Learning the lessons, interrogating the comparisons. In David, K. & Thompson, P.B. *What can nanotechnology learn from biotechnology?* Burlington, MA: Elsevier.
- Marinova, D. et McAleer, M. (2002) Nanotechnology strength indicators: international rankings based on US patents. *Nanotechnology* 14. R1–R7
- Merton, R.K., (1968) The Matthew Effect in Science: The reward and communication systems of science are considered. *Science*. Vol. 159, No 3810. 56-63.

Meyer, M. & Persson, O. (1998). Nanotechnology – Interdisciplinarity, patterns of collaboration and differences in application *Scientometrics*, 42(2), 195-205.

Meyer, M. (2006). Are patenting scientists the better scholars? An exploratory comparison of inventor-authors with their non-inventing peers in nano-science and technology, *Research Policy* 35, 1646-1662.

Meyer, M.S. (2001). Patent citation analysis in a novel field of technology: An exploration of nano-science and nanotechnology. *Scientometrics*, 51(1), 163-183.

Mowery, D.C. (1998). Collaborative R&D: How effective is it?, *Issues in Science and Technology*, 15(1), 37-44.

Mowery, D.C. & Ziedonis, A.A. (2002) Academic patent quality and quantity before and after the Bayh–Dole act in the United States. *Research Policy*. 31-3, 399-418.

Murray, F. & Stern, S. (2004). Do intellectual property right hinder the free flow of scientific knowledge? Evidence from patent-paper pairs.

Murray, F. (2002). Innovation as co-evolution of scientific and technological networks: exploring tissue engineering, *Research Policy* 31, 1389-1403.

Narin, F., Hamilton, K.S. & Olivastro, D. The increasing linkage between U.S. technology and public science. *Research Policy*, 26, 317-330.

Newman, M.E.J. (2001a). Scientific collaboration networks I. Network construction and fundamental results. *Physical Review E*, 64, 016131(1-8).

Newman, M.E.J. (2001b). Scientific collaboration networks II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64, 016132(1-7).

Newman, M.E.J. (2001c). The structure of scientific collaboration networks. *PNAS*, 98(2), 404-409.

OCDE, (1997). Manuel d'Oslo - 2ème édition, Paris

OCDE. (2009). Inventory of national science, technology and innovation policies for nanotechnology 2008, Committee for Scientific and Technological Policy, pp. 1-95.

OECD (2005). A Framework for Biotechnology Statistics. OECD, Paris.



OECD (2005). A Framework for Biotechnology Statistics. OECD, Paris.

OECD, Biotechnology Statistics Database, January 2011; and OECD, Main Science and Technology Indicators Database.

[http://www.oecd.org/document/30/0,3746,en\\_2649\\_34537\\_40146462\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/30/0,3746,en_2649_34537_40146462_1_1_1_1,00.html)

Orsenigo, L., Pammolli, F & Riccaboni, M. (2001). Technological change and network dynamics- lessons from the pharmaceutical industry. *Research Policy* 30, 485-508.

Porter, A. L., Youtie, J., Shapira, P., Shoeneck, D. J. (2007). Refining Search terms for nanotechnology. *Journal of nanoparticle research*. Springer. Vol 10. 715-728.

Porter, M.E. (1998). Clusters and the New Economics of Competition. *Harvard Business Review* – 98609, 77-90.

Rao, S., Ahmad, A., Horsman, W., Kaptein-Russel, P. (2001) The importance of innovation for productivity. Micro-Economic Policy Analysis Branch. Industry Canada.

Rieu, A.-M. (2011). “Fourth helix: sustaining research diversity; the social turn of science and technology policy. The case of Japan” Triple Helix IX International Conference. Stanford, 11-14 July 2011.

Robinson, D.K.R., Rip, A. & Vincent, M. (2007). Technological agglomeration and the emergence of clusters and networks in nanotechnology, *Research Policy* 36, 871-879.

Roco, M.C. (2001), International strategy for nanotechnology research and development, *Journal of Nanoparticle Research* 3, 353-360.

Roco, M.C. (2002). Government Nanotechnology Funding: An International Outlook. *JOM*. September 2002, 22-23.

Rudea, G., Gerdri, P. & Kocaoglu, D. F. (2007). Bibliometrics and social network analysis of the nanotechnology field, *PICMET Proceedings*, 5-9.

Schiffauerova, A. & Beaudry, C. (2011). Gatekeepers of Canadian biotechnology clusters – Where geographical clusters and co-invention networks intersect, *Journal of Engineering and Technology Management* (conditionally accepted 2010-06-29).

Schiffauerova, A., Beaudry, C. (2009). Canadian nanotechnology innovation networks: intra-cluster, inter-cluster and foreign collaboration, *Journal of Innovation Economics* 2(2), 119-146.

Schulte, J. (2005). Growth through Nanotechnology Opportunities and Risks. Nanotechnology: Global Strategies, Industry trends and Applications. John Wiley and Sons Ltd., pp: 97-106.

Shapira, P. & Youtie, J. (2006) Measures for knowledge-based economic development: Introducing data mining techniques to economic developers in the state of Georgia and the US South. Technological Forecasting and Social Change. Vol 73-8, 950-965.

Sparling, D. (2008). A Framework for translating biotechnology experiences to nanotechnology. In David, K. & Thompson, P.B. What can nanotechnology learn from biotechnology? Burlington, MA: Elsevier.

Stephan, P.E., Gurmu, S., Sumell, A.J., Black, G. (2007). Who's patenting in the university? Evidence from the survey of doctorate recipients, Economics of Innovation and New Technology, 16:2, 71-99.

Swanm, P. & Prevezer, M. (1996). A comparison of the dynamics of industrial clustering in computing and biotechnology, Research Policy 25, 1139-1157.

Tchana, J. (2009) Délocalisation des activités de production des grandes entreprises : impact sur les PME. Colloque international. La vulnérabilité des TPE et des PME dans un environnement mondialisé. 27, 28 et 29 mai 2009, INRPME, Trois-Rivières, Canada

Teichert, T. & Mittermayer, M.-A. (2002). "Text Mining for technology monitoring". Engineering Management Conference. IEEE. 596-601. Vol 2.

Termium (2010). Termium Plus : La banque de données terminologiques et linguistiques du Canada. Travaux Publics et Services Gouvernementaux Canada. Consulté le 15-08-11. <http://www.termiumplus.gc.ca/site/termium.php?lang=fra&cont=001>

Terza, J. V., Basu, A., Rathouz P. J., 2008. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. Journal of Health Economics 27(3). 531-543.

Thompson, P.B. (2008). Nano and Bio : How are they alike? How are they different? In David, K. & Thompson, P.B. What can nanotechnology learn from biotechnology? Burlington, MA: Elsevier.

- Toivonen et al. (2005). Data Mining in Gene Mapping. In M.M. Kantardzic & J. Zurada, Next Generation of Data-Mining Applications.(pp.263-294). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Tong, X. and Frame, J. D. (1994). Measuring national technological performance with patent claims data. *Research Policy*, vol. 23, pp. 133–41.
- Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *RAND Journal of Economics*, 21, 172–187.
- Traoré, N. (2004). Biotechnology Use and Development Survey: methodology, issues and responses. Statistics Canada. Science, Innovation and Electronic Information Division.
- UCLA ATS (2011). UCLA Academic Technology Services. [http://www.ats.ucla.edu/stat/mult\\_pkg/faq/general/log\\_transformed\\_regression.htm](http://www.ats.ucla.edu/stat/mult_pkg/faq/general/log_transformed_regression.htm). consulté en novembre 2011.
- Van looy, B., Ranga, M., Callaert, J., Debackere, K. & Zimmermann, E. (2004). Combining entrepreneurial and scientific performance in academia: towards a compounded and reciprocal Matthew-effect, *Research Policy*, 33, 425-441.
- Wasserman, S. & Faust, K. (1994). *Social Network Analysis: Methods and applications*. Cambridge University Press.
- Weiss, S.M., Indurkha, N., Zhang, T., Damerau, F.J. (2005) *Text Mining: Predictive methods for Analyzing Unstructured Information*. New-York: Springer Science + Business Media Inc.
- Wolpert, D.H. & Macready, W.G (1997). “No free lunch theorems for optimization”. *IEEE transactions on evolutionary computation*. Vol 1, No 1. 67-82.
- Wu, P.C. (2005). Need for a new type of venture capital. *Nanotechnology: Global Strategies, Industry trends and Applications*. John Wiley and Sons Ltd., pp:107-125.
- Yoon, B. & Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend, *Management Research* 15, 37-50.
- Youtie, J., Shapira, P. & Porter, A. L. (2008). Nanotechnology publications and citations by leading countries and blocs, *Journal of Nanoparticle Research* 10, 981-986.

Yadolah Dodge (2004) « Valeur p », in Statistique : dictionnaire encyclopédique, Springer, Paris, (2e éd.)

Zucker, L.G. & Darby M. R. (1996). Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry Proceedings of the National Academy of Sciences, 93, 12709-12716.

## ANNEXE A – Tableaux utilisés dans l’analyse des données

Tableau A-1 : Table SIRU

NOM	TYPE	DESCRIPTION
SIRU_ID	INT(11)	Clef primaire d’instance de financement de chercheurs universitaires
SIRU_YEAR	INT(11)	Année de financement
SIRU_SCIENTIST_ID	INT(11)	Numéro d’identification individuel de scientifiques dans SIRU (automatique)
SIRU_SCIENTIST_ORI	VARCHAR(255)	Nom du scientifique recevant le financement
PROJECT TITLE	VARCHAR(255)	Titre du projet pour lequel le scientifique reçoit le financement
SIRU_INSTITUTION	VARCHAR(255)	Numéro d’identification et nom de l’institution auquel le financement est associé
SIRU_INSTITUTION_ORI	VARCHAR(255)	Nom original de l’institution auquel le scientifique est affilié
REGROUPING	VARCHAR(255)	Domaine de recherche
UNIT_NAME	VARCHAR(255)	Nom du département dans lequel le chercheur œuvre
CLUSTER_NAME	VARCHAR(255)	Nom de la grappe à laquelle le chercheur est affilié
COPOU_NAME	VARCHAR(255)	Nom de l’organisation pourvoyeur de fonds
CATEGORY_CODE	INT(11)	Catégorie numéroté du pourvoyeur de fonds
CATEGORY_PROVIDER	VARCHAR(255)	Catégorie du pourvoyeur de fonds
REGROUPING_CATEGORY	VARCHAR(255)	Secteur de financement
PROVIDER_ORIGIN	VARCHAR(255)	Origine géographique (Canadien ou autre)
FINANCING_TYPE	VARCHAR(255)	Type de financement octroyé
FINANCING_TYPE_DETAIL	VARCHAR(255)	Détails du type de financement octroyé
OCTROI	INT(11)	Somme octroyé au scientifique par le pourvoyeur
SCIENTIST_SIRU	VARCHAR(255)	Nom formaté du scientifique recevant les fonds
LAST_NAME_FULL	VARCHAR(255)	Nom de famille du scientifique recevant les fonds
FIRST_NAME_FULL	VARCHAR(255)	Prénom du scientifique recevant les fonds
FIRST_NAME_INITIALS	VARCHAR(255)	Initiaux du prénom du scientifique recevant les fonds
LABEL	VARCHAR(255)	Initiaux du prénom et nom de famille du scientifique recevant les fonds

Tableau A-2 : Données ajoutées à la table SIRU

<b>NOM</b>	<b>TYPE</b>	<b>DESCRIPTION</b>
SCIENTIST_ID_1	VARCHAR(255)	Numéro d'identification du chercheur pour données de brevetage et de financement
BIO	INT(11)	1 si le projet traite de la biotechnologie
NANO	INT(11)	1 si le projet traite de la nanotechnologie
NANO_METROLOGIE&PROCESSUS	INT(11)	1 si le projet traite de la nanométrie ou des nanoprocessus
NANO_DISPOSITIFS&ELECTRONIQUE	INT(11)	1 si le projet traite de la nanodispositifs ou de nanoélectronique
NANO_MEDECINE&BIOTECH	INT(11)	1 si le projet traite de la nanomédecine ou de nanobiotechnologie
NANO_STRUCTURES&MATERIAUX	INT(11)	1 si le projet traite de nanostructures ou de nanomatériaux
ASS_INDIVIDUEL_ID_4	INT(11)	Numéro d'identification associant l'institution pourvoyeur aux cessionnaires de brevets

Tableau A-3 : Information générale sur les brevets

<b>NAME</b>	<b>TYPE</b>	<b>DESCRIPTION</b>
IDPATENT_TITLE_ABSTRACT	VARCHAR(45)	Numéro d'identification individuel de brevet
TITLE	VARCHAR(45)	Titre du brevet
ABSTRACT	TEXT	Résumé du brevet
ANNEE_OBTENTION	INT(11)	Année d'octroi du brevet
ANNEE_APPLICATION	INT(11)	Année d'application du brevet
NB_REVEND	INT(11)	Nombre de revendications associées au brevet
NB_INVENTEURS	INT(11)	Nombre d'inventeurs inscrits sur le brevet
NB_ASSIGNEES	INT(11)	Nombre de cessionnaires inscrits sur le brevet

Tableau A-4 : Information sur le type de brevet

NAME	TYPE	DESCRIPTION
PAT_ID	VARCHAR(255)	Numéro d'identification individuel de brevet
PAT_TYPE	VARCHAR(4)	Catégorie de brevet (nanotech ou biotech)
PAT_DATE	DATETIME	Date d'octrois de brevet
PAT_NBR_REVEN	INT(11)	Nombre de revendications associées aux brevets

Tableau A-5 : Information sur les cessionnaires de brevets

NAME	TYPE	DESCRIPTION
REC_NBR	INT(11)	Individual data identification line
PAT_ID	VARCHAR(255)	Patent identification number
ASS_ID	INT(11)	Assignee identification number
ASS_NAME	VARCHAR(255)	Assignee name as it appears on the patent
ASS_INDIVIDUAL_ID	INT(11)	assignee identification number corrected

Tableau A-6 : Information sur les inventeurs de brevets

NAME	TYPE	DESCRIPTION
REC_NBR	INT(11)	Ligne d'identification individuelle
PAT_ID	VARCHAR(255)	Numéro d'identification de brevets
PAT_DATE	DATETIME	Date d'octroi de brevet
SCIENTIST_ID	INT(11)	Numéro d'identification de chercheur
SCIENTIST_LNAME	VARCHAR(255)	Nom de famille de l'inventeur
SCIENTIST_FNAME	VARCHAR(255)	Prénom de l'inventeur
PAT_SCIENTIST_TOWN	VARCHAR(255)	Ville de l'inventeur sur le brevet
PAT_SCIENTIST_PROV	VARCHAR(255)	Province de l'inventeur sur le brevet
PAT_SCIENTIST_COUNT	VARCHAR(255)	Pays de l'inventeur sur le brevet
PAT_NBR_REVEN	INT(11)	Nombre de revendications associées au brevet
PAT_TYPE	VARCHAR(4)	Type de brevet (nanotech ou biotech)

## ANNEXE B – Tableaux de statistiques descriptives sur le financement et le brevetage

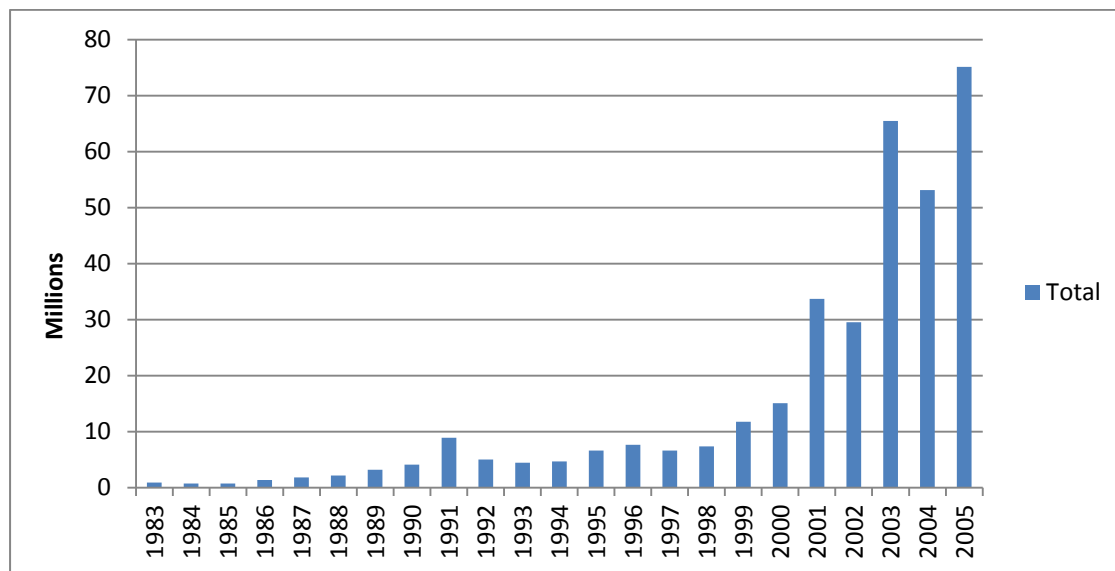
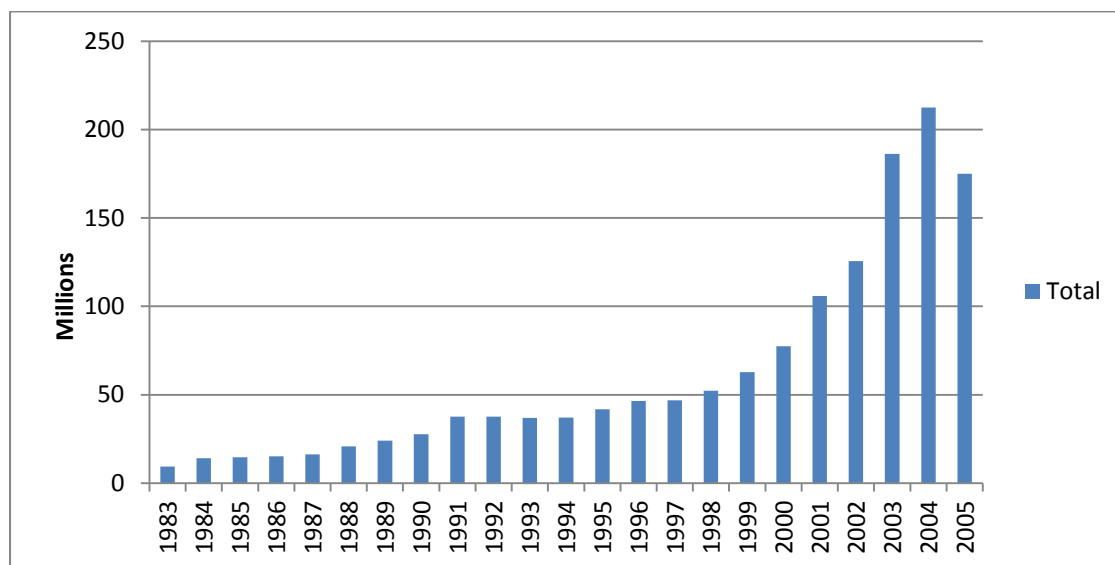


Figure B-1: Financement de chercheurs québécois en nanotechnologie<sup>52</sup>



<sup>52</sup> Ces montants ne reflètent pas les fonds exclusivement en nanotechnologie, mais aussi ceux en nanobiotechnologie. Il en est de même pour les graphs suivants de l'annexe 2.



Figure B-2 : Financement de chercheurs québécois en biotechnologie

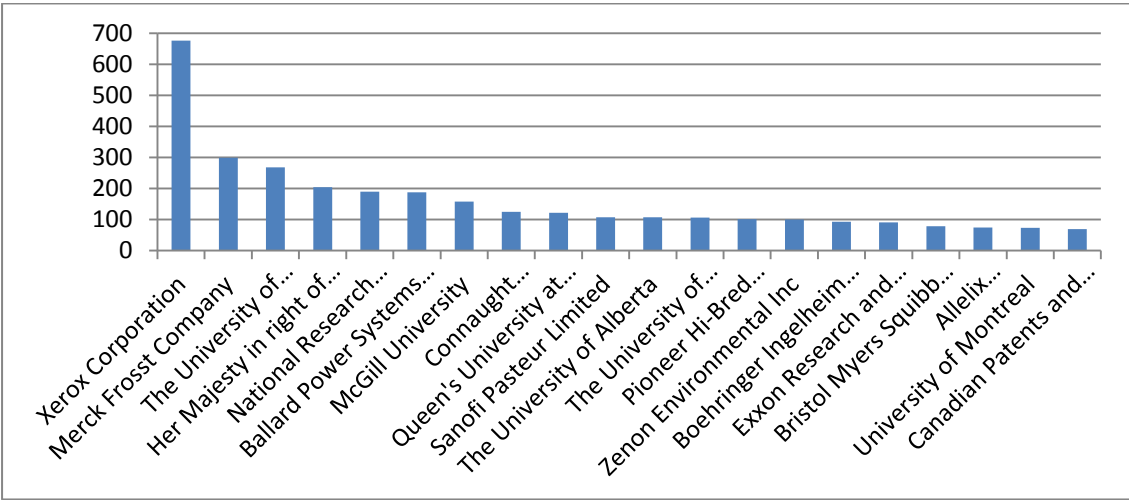


Figure B-3 : Top des détenteurs de brevets en biotechnologie

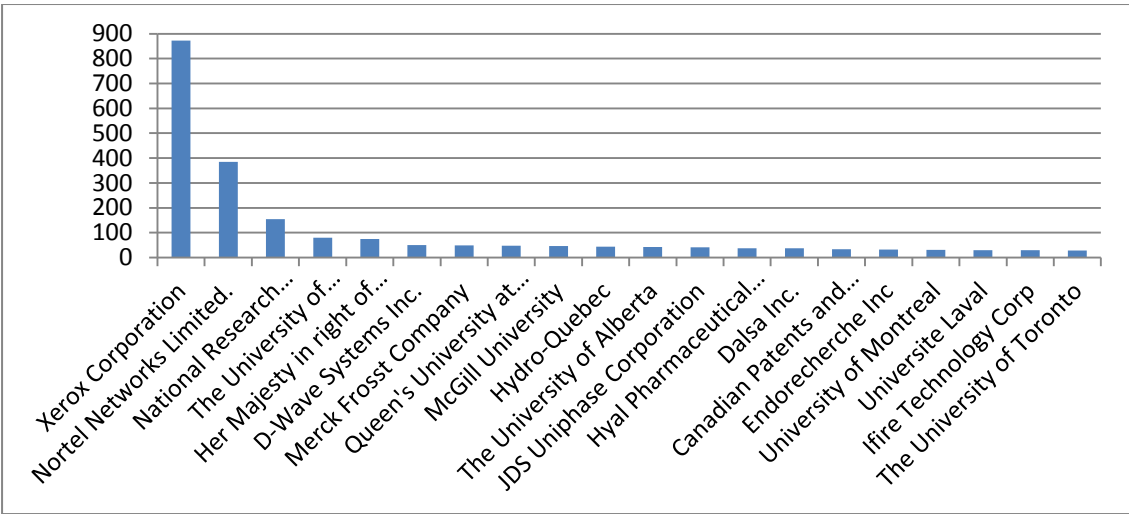


Figure B-4: Top des détenteurs de brevets en nanotechnologie

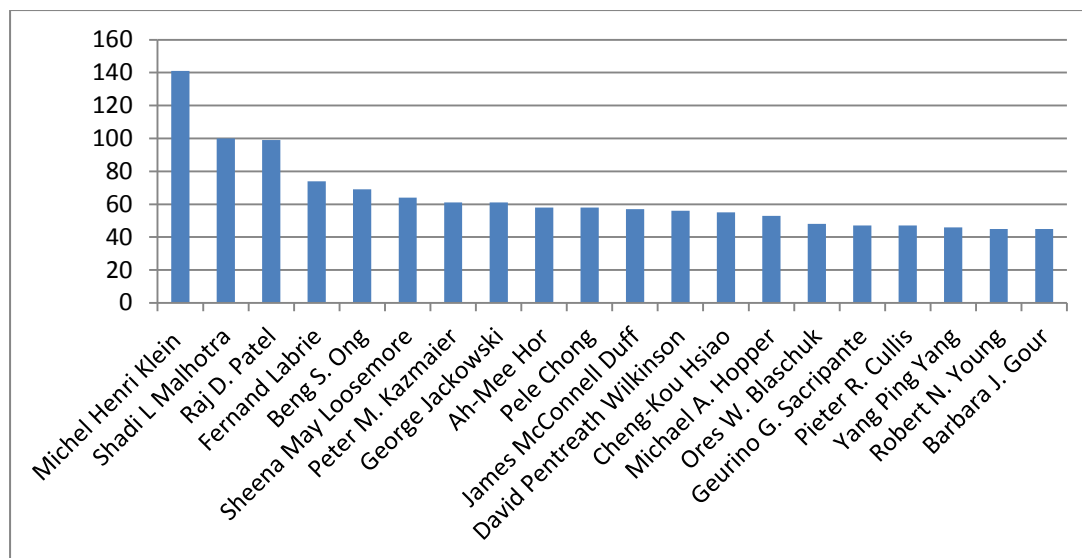


Figure B-5: Meilleurs inventeurs en biotechnologie

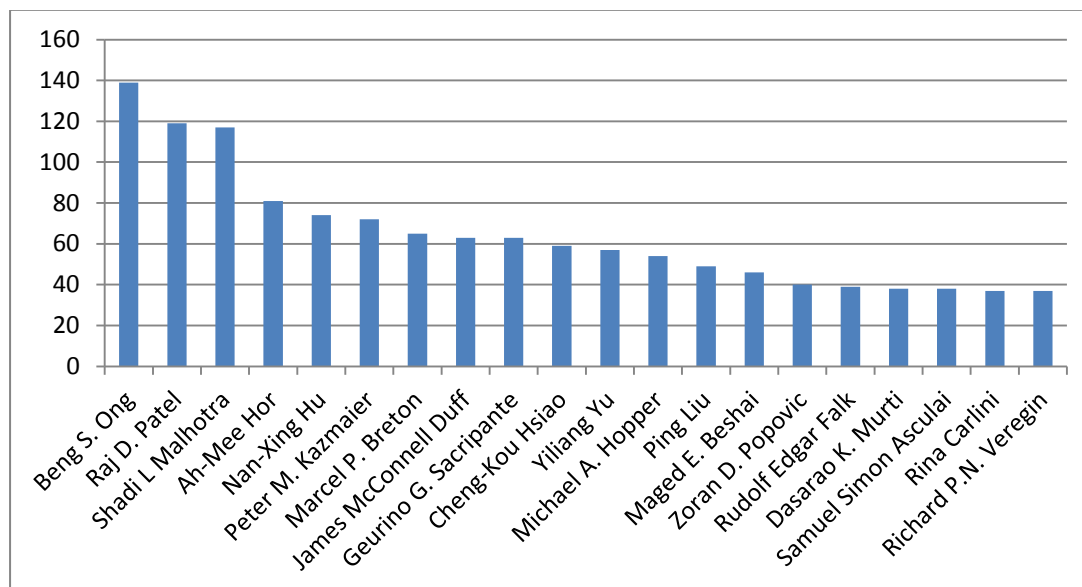


Figure B-6: Meilleurs inventeurs en nanotechnologie

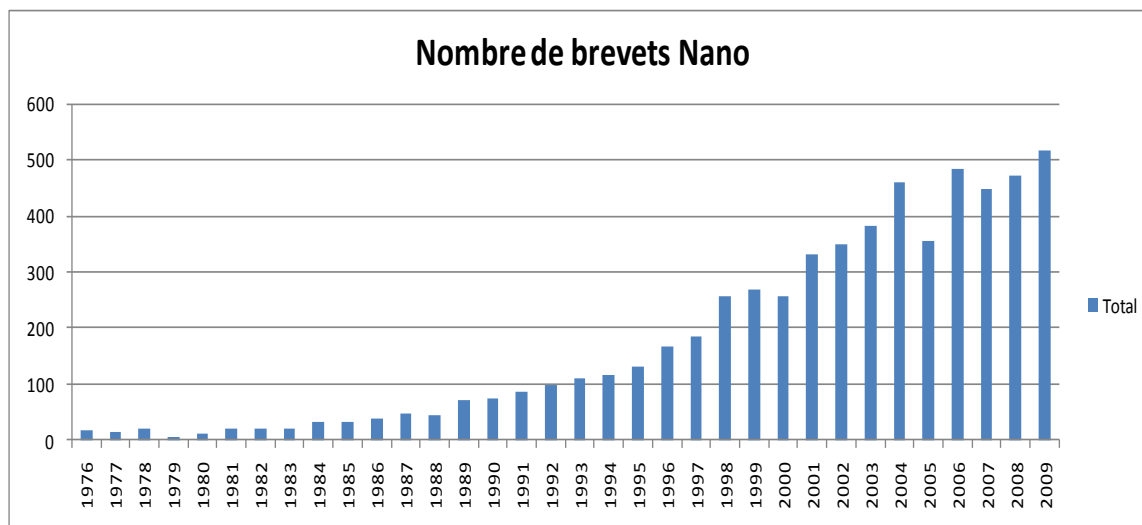


Figure B-7: Évolution du brevetage en nanotechnologie issu d'inventions de chercheurs québécois

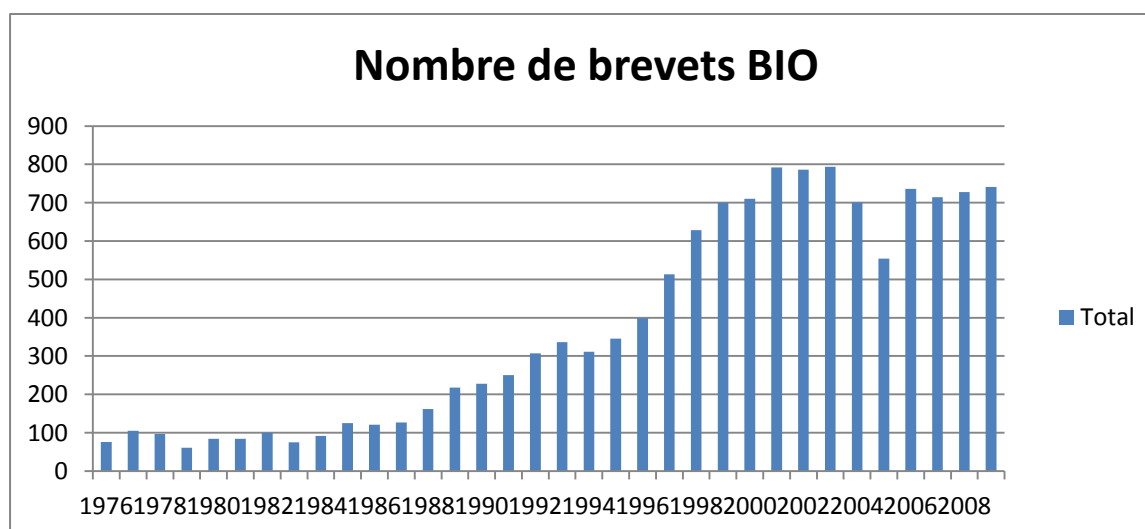


Figure B-8: Évolution du brevetage en biotechnologie issu d'inventions de chercheurs québécois

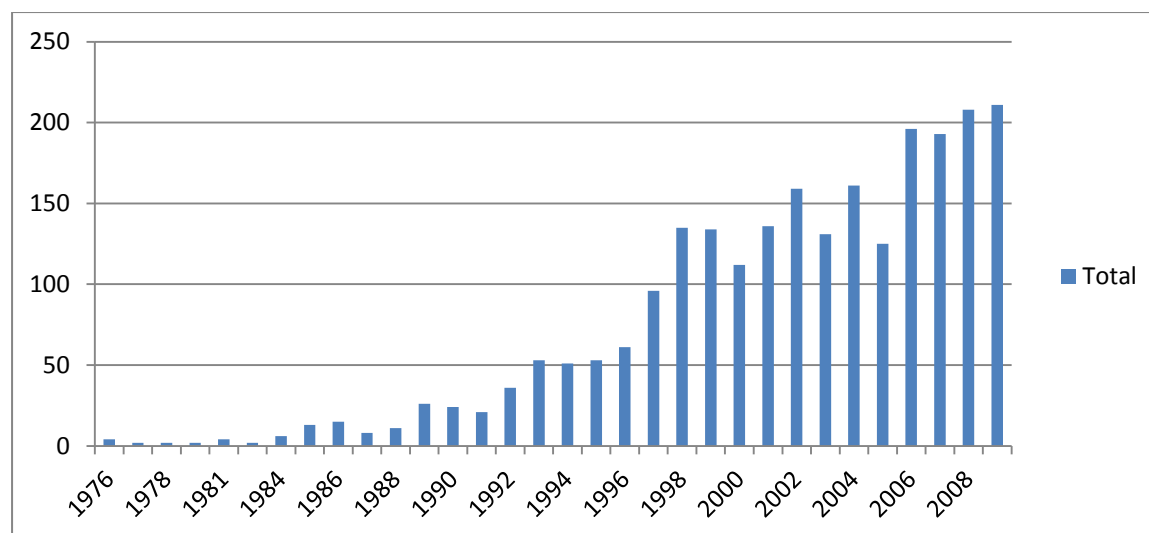


Figure B-9: Évolution du brevetage en nanobiotechnologie issu d'inventions de chercheurs québécois

## ANNEXE C – Extraits de code SQL

```
SELECT siru.ass_individuel_id_4, siru.SCIENTIST_ID_1, siru.siru_year,
patents_scientists.pat_year, patents_assignees.pat_id, Sum(siru.octroi) AS SommeDeoctroi

FROM (siru INNER JOIN patents_scientists ON
siru.SCIENTIST_ID_1=patents_scientists.SCIENTIST_ID) INNER JOIN
patents_assignees ON siru.ass_individuel_id_4=patents_assignees.ASS_Individual_ID

WHERE (((patents_scientists.PAT_ID)=patents_assignees.PAT_ID) And
((siru.siru_year)<=patents_scientists.pat_year))

GROUP BY siru.ass_individuel_id_4, siru.SCIENTIST_ID_1, siru.siru_year,
patents_scientists.pat_year, patents_assignees.pat_id;
```

Extrait du processus de bouclage.

```
SELECT DISTINCT * FROM `core_nano_bio`.`patents_assignees` WHERE
`core_nano_bio`.`patents_assignees`.`pat_id` IN (SELECT DISTINCT
`core_nano_bio`.`patent_type`.`pat_id` FROM `core_nano_bio`.`patent_type` WHERE
`core_nano_bio`.`patent_type`.`pat_type`="BIO");
```

Extrait pour la sélection du nombre individuel de cessionnaires de brevets en biotechnologie

```

SELECT DISTINCT `core_nano_bio`.`patents_scientists`.`pat_id`,
`core_nano_bio`.`patents_scientists`.`scientist_ID` FROM
`core_nano_bio`.`patents_scientists` WHERE
`core_nano_bio`.`patents_scientists`.`pat_id` IN (SELECT DISTINCT
`core_nano_bio`.`patent_type`.`pat_id` from `core_nano_bio`.`patent_type` WHERE
`core_nano_bio`.`patent_type`.`pat_type`="BIO") AND
`core_nano_bio`.`patents_scientists`.`scientist_id` IN (SELECT DISTINCT
`core_nano_bio`.`siru_complet`.`scientist_id_1` FROM
`core_nano_bio`.`siru_complet`);

```

Extrait de chercheurs universitaires qui sont mentionnés comme inventeurs sur des brevets en biotech

```

SELECT AVG(`core_nano_bio`.`patent_info`.`nb_revend`) FROM
`core_nano_bio`.`patent_info` WHERE
`core_nano_bio`.`patent_info`.`idpatent_title_abstract` IN (SELECT DISTINCT
`core_nano_bio`.`x_rami_nano_loop_pat`.`pat_id` FROM
`core_nano_bio`.`x_rami_nano_loop_pat`);

```

Extrait pour déterminer le nombre moyen de revendications des brevets inclus en boucle en nanotech

```

SELECT AVG(`core_nano_bio`.`patent_info`.`nb_revend`) FROM
`core_nano_bio`.`patent_info` WHERE
(`core_nano_bio`.`patent_info`.`idpatent_title_abstract` IN (SELECT DISTINCT
`core_nano_bio`.`patent_type`.`pat_id` from `core_nano_bio`.`patent_type` WHERE
`core_nano_bio`.`patent_type`.`pat_type`="BIO") AND
`core_nano_bio`.`patent_info`.`idpatent_title_abstract` NOT IN (SELECT DISTINCT
`core_nano_bio`.`x_rami_bio_loop_patents`.`pat_id` FROM
`core_nano_bio`.`x_rami_bio_loop_patents`));

```

Extrait pour déterminer la moyenne du nombre de revendications de brevets en biotech exclus de boucles

```
SELECT * FROM `core_nano_bio`.`citing_patents_us` WHERE
(`core_nano_bio`.`citing_patents_us`.`pat_id_cited` IN (SELECT DISTINCT
`core_nano_bio`.`patent_type`.`pat_id` FROM `core_nano_bio`.`patent_type`
WHERE `core_nano_bio`.`patent_type`.`pat_type`="BIO") AND
`core_nano_bio`.`citing_patents_us`.`pat_id_cited` NOT IN (SELECT DISTINCT
`core_nano_bio`.`x_ami_bio_loop_patents`.`pat_id` FROM
`core_nano_bio`.`x_ami_bio_loop_patents`));
```

Extrait pour déterminer la moyenne des citations de brevets biotech exclus de boucles

```
SELECT AVG(`core_nano_bio`.`patent_info`.`nb_revend`) FROM
`core_nano_bio`.`patent_info` WHERE
`core_nano_bio`.`patent_info`.`idpatent_title_abstract` IN (SELECT DISTINCT
`core_nano_bio`.`x_ami_nano_loop_pat`.`pat_id` FROM
`core_nano_bio`.`x_ami_nano_loop_pat` WHERE
`core_nano_bio`.`x_ami_nano_loop_pat`.`firm owned`=1);
```

SQL excerpt for determining average nb of claims for nano patents included in loops & belonging to firms

```
SELECT `core_nano_bio`.`patents_scientists`.`pat_id`,
`core_nano_bio`.`patents_scientists`.`scientist_id` FROM
`core_nano_bio`.`patents_scientists` WHERE
(`core_nano_bio`.`patents_scientists`.`scientist_id` IN (SELECT
`core_nano_bio`.`octrois_scientist_org`.`scientist_id` FROM
`core_nano_bio`.`octrois_scientist_org`) OR
`core_nano_bio`.`patents_scientists`.`scientist_id` IN (SELECT DISTINCT
`core_nano_bio`.`x_ami_siru2`.`scientist_id_1` FROM
`core_nano_bio`.`x_ami_siru2`));
```

Selection d'inventeurs universitaires

```

SELECT pat_merge11.scientist_id, pat_merge11.annee_application,
count(pat_merge11.pat_id) AS nb_patents, sum(pat_merge11.nb_revend) AS
sum_nb_claims, sum (pat_merge11.nb_assignees) AS sum_assignees,
sum(pat_merge11.nb_inventeurs) AS sum_inventors,
sum(pat_merge11.pat_cit_3_yrs) AS sum_pat_cit_3_yrs,
sum(pat_merge11.pat_cit_5_yrs) AS sum_pat_cit_5_yrs,
sum(pat_merge11.pat_cit_7_yrs) AS sum_pat_cit_7_yrs INTO pat_merge22
FROM pat_merge11
GROUP BY pat_merge11.scientist_id, pat_merge11.annee_application;

```

Une des étapes d'insertion et de consolidation de données de brevetage dans la matrice

```

select `core_nano_bio`.`patents_scientists`.`pat_id`,
`core_nano_bio`.`patent_info`.`annee_application` from
`core_nano_bio`.`patents_scientists` inner join `core_nano_bio`.`patent_info` on
`core_nano_bio`.`patents_scientists`.`pat_id` =
`core_nano_bio`.`patent_info`.`idpatent_title_abstract` where
`core_nano_bio`.`patents_scientists`.`scientist_lname`="Labrie" and
`core_nano_bio`.`patent_info`.`annee_application`='1995';

```

Une étape de vérification aléatoire permettant de déceler des erreurs dans les données utilisées lors de l'analyse, plus spécifiquement pour le nombre d'application de brevets à l'année 1995.

```

select * from `core_nano_bio`.`citing_patents_us` where
`core_nano_bio`.`citing_patents_us`.`pat_id_cited`="5434146" and
`core_nano_bio`.`citing_patents_us`.`pat_id_cited_year_granted`-
`core_nano_bio`.`citing_patents_us`.`pat_id_citing_year`<=7;

```

Une autre étape de verification, celle-ci identifier les instances où le brevet 5434146 est cité dans les 7 ans suivants sa date d'octroi.



## ANNEXE D – Inconvénients de l'utilisation de forage de données et de l'inspection manuelle pour la validation des boucles d'innovation

Tableau D-1: Inconvénients d'utiliser le forage de données dans l'évaluation des boucles d'innovation

	Exemple		
Problèmes	Loop ID	Titre de Projet	Titre de Brevets
Titre de projet en français et titre de brevets en anglais	256	Développement de nouveaux matériaux de stockage de l'hydrogène à base d'hydrures métalliques nanocristallins.	Method of rapidly carrying out a hydrogenation of a hydrogen storage material
	479	ETUDE DE FAISABILITE POUR LA PRODUCTION DE POUDRES ULTRA- FINES D'ARGENT	Method and transferred arc plasma system for production of fine and ultrafine powders
	729	Etude des composés ayant des propriétés antivirales contre différents virus Herpes	Agents with leukotriene B4-like activity as antiherpes viral agents
No relevance	982	RAPID IDENTIFICATION & MAPPING OF STI'S	Methods of treating osteoarthritis with inducible nitric oxide synthase inhibitors
Different patent/fund but 2 similar patents	242	THE EFFECTS OF ESTROGENS ON E-CADHERIN GENE EXPRESSION	Compounds and methods for modulating nonclassical cadherin-mediated functions
	247	THE EFFECTS OF ESTROGENS ON E-CADHERIN GENE EXPRESSION	Compounds and methods for modulating adhesion molecule function

Tableau D-2: Désavantages d'effectuer une révision individuelle des boucles d'innovation

Problèmes	Exemple		
	Loop ID	Project Title	Patent Title
Détection d'erreur	928	AIDE A LA CREATION LITTERAIRE ET MUSICALE	Integration cassette for improvement of transgenesis in eukaryotes
	1058	EVOLUTION RECENTE ET FUTURE DE L'HABITAT ESTIVAL DES CARIBOUS DE LA RIVIERE GEO RGE	Method of rapidly carrying out a hydrogenation of a hydrogen storage material
Impossible de déterminer si l'argent a mené au brevet	1136	RIBOSOME STUDIES*	Catalytic DNA
	1279	PROJET DE COOPERATION ENTRE: IREQ ET MCGILL	Method for inducing hydrogen desorption from a metal hydride
	1285	RIBOSOME STUDIES*	Catalytic DNA
	1313	CONTRIBUTION DE IAF BIOCHEM A LA CHAIRE INDUS- TRIELLE ET A LA DIVISION DES PRODUITS BIOCHIMIQUES.	2-substituted-5-substituted-1,3-oxathiolanes with antiviral properties
	1622	Simulateur dynamique in vitro du tractus gastro-intestinal	Method and system for modulation of microbial cell characteristics
	1687	SOLID SUPRAMOLECULAR NETWORKS AND THEIR APPLICATION TO DIRECT MARKING TECHNOLOGY (MOLECULAR TECTONICS) U.I.\$*	Alkylated tetrakis(triaminotriazine) compounds and phase change inks containing same
	1789	DESIGN ET ETUDE DES PERFORMANCES D'UNITES DE LA TECHNOLOGIE OXYJET.	Process for producing low DP microcrystalline cellulose
	1871	SYNTHÈSE D'ACIDES CARBOXYLIQUES ET SYNTHÈSE DE COLORANTS DIBROMÉS*	GH secretagogues and uses thereof

## ANNEXE E – Graphes supplémentaires

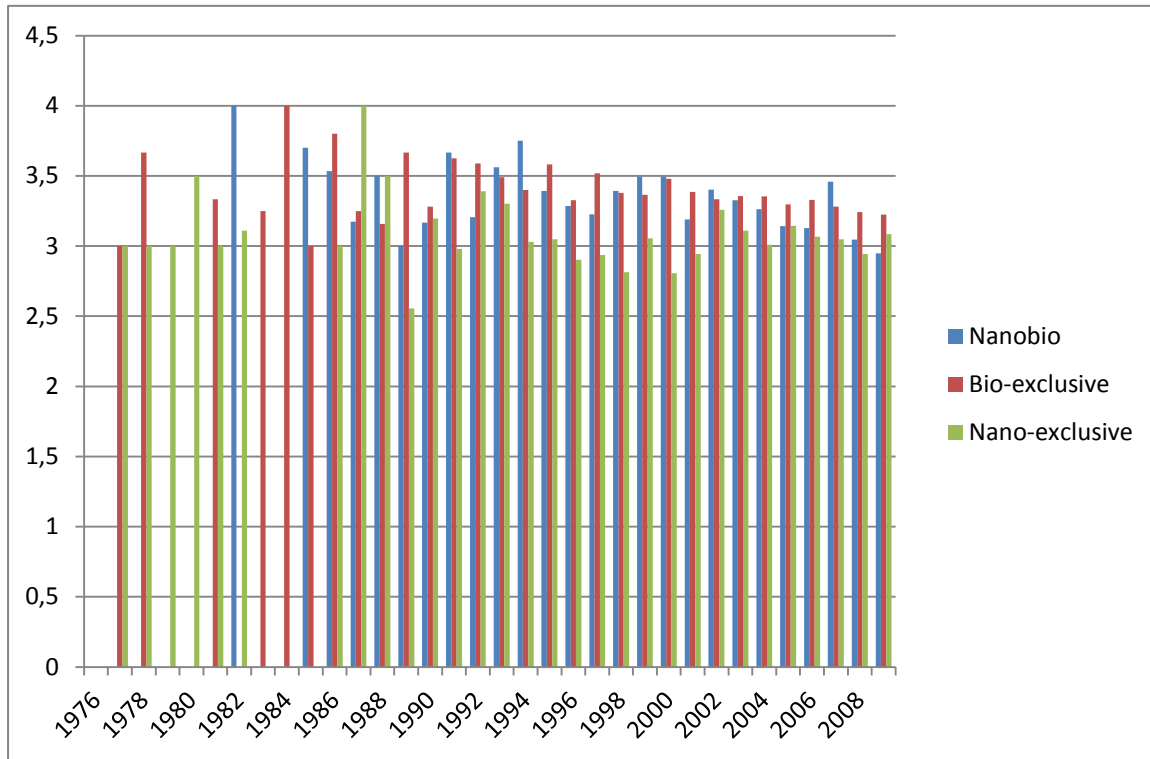


Figure E-1: Average degree of application of knowledge associated to patented inventions

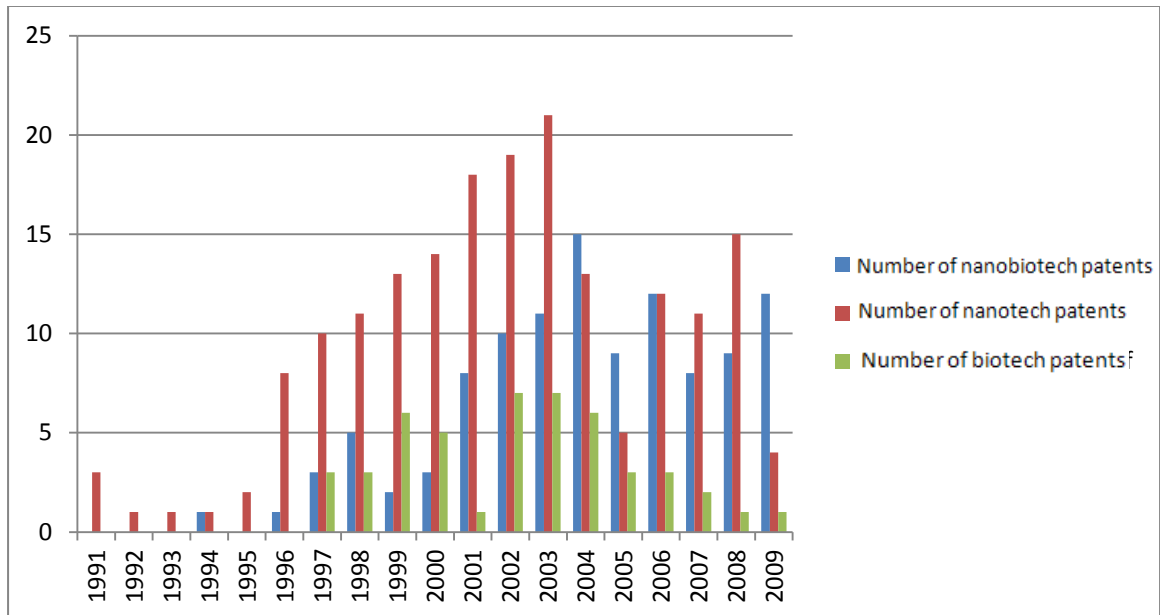


Figure E-2: Number of patents derived from loops

## ANNEXE F – Procédure pour tester différences de moyennes et variances

### COMPARAISONS GÉNÉRALES DE QUALITÉ ET DÉGRÉE D'APPLICATION.

Nombre de revendications :

- nanotechnologie(1) versus biotechnologie(2)
- nanotechnologie(1) versus nanobiotechnologie(3)
- biotechnologie(2) versus nanobiotechnologie(3)

Nombre de citations :

- nanotechnologie(1) versus biotechnologie(2)
- nanotechnologie(1) versus nanobiotechnologie(3)
- *biotechnologie(2) versus nanobiotechnologie(3)*<sup>53</sup>

Degré d'application

- nanotechnologie(1) versus biotechnologie(2)
- nanotechnologie(1) versus nanobiotechnologie(3)
- biotechnologie(2) versus nanobiotechnologie(3)

---

<sup>53</sup> À titre d'exemple, le détail de cette comparaison se trouve ci-dessous et les autres se trouvent à l'Annexe 5.

**COMPARAISONS DE BOUCLES D'INNOVATION QUÉBÉCOISES (dans et hors-boucle)**  
**DE QUALITÉ ET DEGRÉ D'APPLICATION.**

Nombre de revendications

- Nanotechnologie en boucle versus hors boucle
- Biotechnologie en boucle versus hors boucle
- Nanobiotechnologie en boucle versus hors boucle

Nombre de citations

- Nanotechnologie en boucle versus hors boucle
- Biotechnologie en boucle versus hors boucle
- Nanobiotechnologie en boucle versus hors boucle

Degré d'application

- nanotechnologie selon la position par rapport à la moyenne
- nanotechnologie selon la position par rapport à la médiane
- nanotechnologie selon les quartiles 1 ou 4 (boucles)
- biotechnologie selon la position par rapport à la moyenne
- biotechnologie selon la position par rapport à la médiane
- biotechnologie selon les quartiles 1 ou 4 (boucles)
- nanobiotechnologie selon la position par rapport à la moyenne
- nanobiotechnologie selon la position par rapport à la médiane
- nanobiotechnologie selon les quartiles 1 ou 4 (boucles)

Table F-1 : Variables compilées

Variable	Obs	Mean	Std. Dev.	Min	Max
pat_id	16418	8209.5	4739.613	1	16418
idpatent	0				
bio	16418	.7821294	.4128116	0	1
nano	16418	.3642953	.4812466	0	1
year_grant~g	16323	1999.538	7.562925	1976	2009
year_appli~n	16323	1996.803	7.126726	1944	2009
nb_claims	16323	18.68884	14.94961	1	249
nb_inventors	16323	2.941371	2.06616	1	32
nb_univ_in~s	16417	.3656576	.6905427	0	12
prop_univ_~v	16417	.1271489	.2484626	0	1
with_acad_~v	16417	.2763599	.4472105	0	1
wo_acad_inv	16417	.7236401	.4472105	0	1
nb_assignees	16323	.9506831	.4432851	0	10
nb_cit_3_yrs	16418	1.425204	2.845182	0	82
nb_cit_5_yrs	16418	2.824096	5.234723	0	133
nb_cit_7_yrs	16418	4.13942	7.611908	0	171
of_loop	16418	.0206481	.1422074	0	1
coll_type_~m	16418	17.653	12.09205	0	61
coll_type_~l	16418	11.77738	7.882139	0	33
deg_applic~n	3451	3.2585	.700284	0	4
type_patent	16418	1.928554	.5993435	1	3
qc_researc~r	16418	.0699233	.2550254	0	1
qc_researc~p	16418	.0905713	.3516701	0	2
chercheur_~p	275	1.636364	.4819227	1	2



## Nombre de revendications : nanotechnologie(1) versus biotechnologie(2)

### Échantillon 1

```
. sdtest nb_claims, by(type_patent)
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	3555	21.48748	.2531614	15.09445	20.99113	21.98384
2	3457	17.1562	.249946	14.69589	16.66615	17.64626
combined	7012	19.35211	.1797852	15.0548	18.99968	19.70454

ratio = sd(1) / sd(2) f = 1.0550  
 Ho: ratio = 1 degrees of freedom = 3554, 3456

Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1  
 Pr(F < f) = 0.9433 2\*Pr(F > f) = 0.1133 Pr(F > f) = 0.0567

Parce que nous ne pouvons rejeter l'hypothèse selon laquelle il y a équivalence des variances au seuil de significativité de 5%, nous pouvons procéder avec un test ANOVA.

```
. oneway nb_claims type_patent, tabulate
```

type_patent	Summary of Nb_claims		
	Mean	Std. Dev.	Freq.
1	21.487482	15.094452	3555
2	17.156205	14.695892	3457
Total	19.352111	15.054798	7012

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	32879.7964	1	32879.7964	148.11	0.0000
Within groups	1556141.84	7010	221.988851		
Total	1589021.64	7011	226.646932		

Bartlett's test for equal variances:  $\chi^2(1) = 2.5080$  Prob> $\chi^2 = 0.113$

Le test Bartlett's nous permet de décider s'il est possible se fier au test F. Dans ce cas-ci, nous pouvons dire qu'il est possible de le faire car prob >  $\chi^2$  se trouve à être plus grande que le seuil de 5%. Parce que Prob > F est inférieure à 5%, il est possible pour nous de rejeter l'hypothèse nulle selon laquelle il n'y a aucune différence entre le nombre de revendications des brevets portant exclusivement sur la nanotechnologie et celles portant exclusivement sur la biotechnologie.



## Échantillon 2

```
. sdtest nb_claims, by(type_patent)
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	3555	21.48748	.2531614	15.09445	20.99113	21.98384
2	3477	17.44751	.2507943	14.78836	16.95579	17.93923
combined	7032	19.4899	.1798149	15.07874	19.13741	19.84239

ratio = sd(1) / sd(2) f = 1.0418  
 Ho: ratio = 1 degrees of freedom = 3554, 3476

Ha: ratio < 1      Ha: ratio != 1      Ha: ratio > 1  
 Pr(F < f) = 0.8877      2\*Pr(F > f) = 0.2247      Pr(F > f) = 0.1123

Parce que nous ne pouvons rejeter l'hypothèse selon laquelle il y a équivalence des moyennes au seuil de significativité de 5%, nous pouvons procéder avec un test ANOVA.

```
. oneway nb_claims type_patent, tabulate
```

type_patent	Summary of Nb_claims		
	Mean	Std. Dev.	Freq.
1	21.487482	15.094452	3555
2	17.447512	14.788358	3477
Total	19.489903	15.07874	7032

Source	Analysis of Variance			F	Prob > F
	SS	df	MS		
Between groups	28689.4192	1	28689.4192	128.47	0.0000
Within groups	1569937.86	7030	223.319753		
Total	1598627.28	7031	227.368409		

Bartlett's test for equal variances: chi2(1) = 1.4746 Prob>chi2 = 0.225

Il est possible de se fier au test F car la prob >chi2 se trouve à être plus grande que le seuil de 5%. Parce que Prob >F est inférieur à 5%, il est possible pour nous de rejeter l'hypothèse nulle selon laquelle il n'y a aucune différence entre le nombre de revendications des brevets portant exclusivement sur la nanotechnologie et celles portant exclusivement sur la biotechnologie.

## Échantillon 3

```
. sdtest nb_claims, by(type_patent)
```

```
Variance ratio test
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	3555	21.48748	.2531614	15.09445	20.99113	21.98384
2	3448	16.94838	.2315293	13.59532	16.49443	17.40232
combined	7003	19.25261	.1739033	14.55291	18.9117	19.59351

```
ratio = sd(1) / sd(2)                                f = 1.2327
Ho: ratio = 1                                         degrees of freedom = 3554, 3447
```

```
Ha: ratio < 1                                Ha: ratio != 1                                Ha: ratio > 1
Pr(F < f) = 1.0000                        2*Pr(F > f) = 0.0000                        Pr(F > f) = 0.0000
```

Parce que nous ne pouvons accepter l'hypothèse selon laquelle il y a équivalence des variances au seuil de significativité de 5%, nous devons effectuer un test Kruskal-Wallis (kwallis) dans Stata, qui est à toutes fins pratiques une version non-paramétrique du test ANOVA.

```
. kwallis nb_claims, by(type_patent)
```

```
Kruskal-Wallis equality-of-populations rank test
```

type_p~t	Obs	Rank Sum
1	3555	1.38e+07
2	3448	1.07e+07

```
chi-squared = 249.387 with 1 d.f.
probability = 0.0001
```

```
chi-squared with ties = 249.611 with 1 d.f.
probability = 0.0001
```

En regardant les probabilités Chi2 avec et sans liens, nous pouvons affirmer qu'il y a une différence statistiquement significative entre le nombre de citations de brevets portant exclusivement sur la nanotechnologie et ceux portant exclusivement sur la biotechnologie. Parce que nous sommes dans une situation où ne prenons pas pour acquis qu'il y a une distribution normale de la variable du nombre de revendications, nous effectuons un test de Mann-Whitney.

```
. ranksum nb_claims, by(type_patent)
Two-sample Wilcoxon rank-sum (Mann-Whitney) test
```

type_patent	obs	rank sum	expected
1	3555	13785354	12449610
2	3448	10739152	12074896
combined	7003	24524506	24524506

```
unadjusted variance 7.154e+09
adjustment for ties -6413188.6
adjusted variance 7.148e+09
Ho: nb_claims(type_patent==1) = nb_claims(type_patent==2)
z = 15.799
Prob > |z| = 0.0000
```

Il est clair qu'il y a une différence significative entre les deux moyennes du nombre de revendications.

Donc nous avons vu ici que les moyennes du nombre de revendications dans le cas des 3 échantillons sont significativement différentes

### Nombre de revendications : nanotechnologie (1) versus nanobiotechnologie (3)

Nous procédons avec un test de déviation standard (sdtest) pour évaluer la similarité des variances de chaque catégorie que nous désirons examiner.

```
. sdtest nb_claims, by (type_patent)
Variance ratio test
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
1	3555	21.48748	.2531614	15.09445	20.99113 21.98384
3	2386	21.0637	.3325037	16.24171	20.41168 21.71573
combined	5941	21.31729	.2019439	15.56541	20.9214 21.71317

```
ratio = sd(1) / sd(3) f = 0.8637
Ho: ratio = 1 degrees of freedom = 3554, 2385
Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1
Pr(F < f) = 0.0000 2*Pr(F < f) = 0.0001 Pr(F > f) = 1.0000
```

Nous n'avons pas besoin de décomposer en sous-échantillons car le nombre d'observations de chaque élément à comparer sont dans le même ordre de grandeur. Parce que nous ne pouvons accepter l'hypothèse selon laquelle il y a équivalence des variances au seuil de significativité de 5%, nous devons effectuer un test Kruskal-Wallis (kwallis).

```
. kwallis nb_claims, by (type_patent)
Kruskal-Wallis equality-of-populations rank test
```

type_p~t	Obs	Rank Sum
1	3555	1.07e+07
3	2386	6.95e+06

```
chi-squared =      4.573 with 1 d.f.
probability =      0.0325
```

```
chi-squared with ties =      4.577 with 1 d.f.
probability =      0.0324
```

Nous pouvons affirmer qu'il y a une différence statistiquement significative entre le nombre de citations de brevets exclusivement nanotechnologiques et ceux exclusivement nanobiotechnologiques. Nous effectuons alors un test de Mann-Whitney (ranksum).

```
. ranksum nb_claims, by (type_patent)
Two-sample wilcoxon rank-sum (Mann-Whitney) test
```

type_patent	obs	rank sum	expected
1	3555	10700486	10561905
3	2386	6950225.5	7088806
combined	5941	17650711	17650711

```
unadjusted variance  4.200e+09
adjustment for ties  -3254033.4
adjusted variance    4.197e+09
```

```
Ho: nb_cla~s(type_p~t==1) = nb_cla~s(type_p~t==3)
      z =      2.139
      Prob > |z| =      0.0324
```

Il y a une différence statistiquement significative entre les deux moyennes.

## Nombre de revendications : biotechnologie(2) versus nanobiotechnologie(3)

### Échantillon 1

```
. sdtest nb_claims, by(type_patent)
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
2	2592	17.13156	.2739547	13.9475	16.59437	17.66875
3	2386	21.0637	.3325037	16.24171	20.41168	21.71573
combined	4978	19.01627	.2156693	15.21654	18.59346	19.43908

```

ratio = sd(2) / sd(3)                                f = 0.7374
Ho: ratio = 1                                         degrees of freedom = 2591, 2385

Ha: ratio < 1                Ha: ratio != 1                Ha: ratio > 1
Pr(F < f) = 0.0000          2*Pr(F < f) = 0.0000          Pr(F > f) = 1.0000

```

Il y a non-égalité des variances, nous effectuons donc initialement un test Kruskal-Wallis.

```
. kwallis nb_claims, by(type_patent)
```

Kruskal-Wallis equality-of-populations rank test

type_p~t	Obs	Rank Sum
2	2592	5.94e+06
3	2386	6.45e+06

```

chi-squared = 100.930 with 1 d.f.
probability = 0.0001

```

```

chi-squared with ties = 101.018 with 1 d.f.
probability = 0.0001

```

Il y a une différence statistiquement significative entre les deux populations

```
. ranksum nb_claims, by(type_patent)
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

type_patent	obs	rank sum	expected
2	2592	5943869.5	6452784
3	2386	6448861.5	5939947
combined	4978	12392731	12392731

```

unadjusted variance 2.566e+09
adjustment for ties -2218063.2
-----
adjusted variance 2.564e+09

```

```

Ho: nb_cla~s(type_p~t==2) = nb_cla~s(type_p~t==3)
z = -10.051
Prob > |z| = 0.0000

```

Afin de confirmer la différence, nous effectuons un test de Mann-Whitney et trouvons effectivement qu'il y a une distinction significative entre les deux moyennes de biotechnologie et nanobiotechnologie.

## Échantillon 2

```
. sktest nb_claims, noadjust
```

```

              Skewness/Kurtosis tests for Normality
-----+-----+-----+-----+-----+
Variable | Pr(Skewness)  Pr(Kurtosis)  chi2(2)  joint Prob>chi2
-----+-----+-----+-----+-----+
nb_claims |      0.000      0.000    3352.75    0.0000

```

Nous rejetons la normalité des données puisque la prob>chi2 est inférieure à 0.05.

```
. sdtest nb_claims, by(type_patent)
```

Variance ratio test

```

-----+-----+-----+-----+-----+
Group | Obs   Mean   Std. Err.   Std. Dev.   [95% Conf. Interval]
-----+-----+-----+-----+-----+
  2 | 2625  17.00343   .2844996   14.57627    16.44556    17.56129
  3 | 2386  21.0637   .3325037   16.24171    20.41168    21.71573
-----+-----+-----+-----+-----+
combined | 5011  18.93674   .2192909   15.52326    18.50683    19.36665
-----+-----+-----+-----+-----+
ratio = sd(2) / sd(3)                                f = 0.8054
Ho: ratio = 1                                         degrees of freedom = 2624, 2385

Ha: ratio < 1          Ha: ratio != 1          Ha: ratio > 1
Pr(F < f) = 0.0000    2*Pr(F < f) = 0.0000    Pr(F > f) = 1.0000

```

En effectuant un test de déviation standard, nous remarquons qu'il y a non-égalité des variances.

Nous laissons donc le test ANOVA de côté et procédons avec un test Kruskal-Wallis.

```
. kwallis nb_claims, by(type_patent)
```

Kruskal-Wallis equality-of-populations rank test

```

+-----+-----+-----+
| type_p~t | Obs | Rank Sum |
+-----+-----+-----+
|      2 | 2625 | 6.02e+06 |
|      3 | 2386 | 6.53e+06 |
+-----+-----+-----+

```

```
chi-squared = 117.324 with 1 d.f.
probability = 0.0001
```

```
chi-squared with ties = 117.428 with 1 d.f.
probability = 0.0001
```

Nous identifions une différence statistiquement significative entre les deux populations.

```
. ranksum nb_claims, by(type_patent)
Two-sample Wilcoxon rank-sum (Mann-Whitney) test
```

type_patent	obs	rank sum	expected
2	2625	6024251.5	6578250
3	2386	6533314.5	5979316
combined	5011	12557566	12557566

```
unadjusted variance 2.616e+09
adjustment for ties -2297238.9
adjusted variance 2.614e+09

Ho: nb_cla~s(type_p~t==2) = nb_cla~s(type_p~t==3)
      z = -10.836
      Prob > |z| = 0.0000
```

Le test de Mann-Whitney permet de confirmer qu'il y a une différence statistiquement significative des deux populations de l'échantillon 2.

### Échantillon 3

```
. sktest nb_claims, noadjust
```

Skewness/Kurtosis tests for Normality

Variable	Pr(Skewness)	Pr(Kurtosis)	chi2(2)	joint Prob>chi2
nb_claims	0.000	0.000	2810.00	0.0000

Une fois de plus, nous rejetons la normalité des données.

```
. sdtest nb_claims, by(type_patent)
Variance ratio test
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
2	2603	17.48483	.2856791	14.57523	16.92464 18.04501
3	2386	21.0637	.3325037	16.24171	20.41168 21.71573
combined	4989	19.19643	.2193976	15.49668	18.76632 19.62655

```
ratio = sd(2) / sd(3)
Ho: ratio = 1
Ha: ratio < 1
Pr(F < f) = 0.0000
Ha: ratio != 1
2*Pr(F < f) = 0.0000
Ha: ratio > 1
Pr(F > f) = 1.0000
f = 0.8053
degrees of freedom = 2602, 2385
```

Nous effectuons donc un sdtest et trouvons que nous pouvons rejeter l'hypothèse d'égalité des variances. Il est donc nécessaire de faire un test de kwallis.

```
. kwallis nb_claims, by(type_patent)
Kruskal-Wallis equality-of-populations rank test
```

type_p~t	Obs	Rank Sum
2	2603	6.01e+06
3	2386	6.44e+06

```
chi-squared = 91.138 with 1 d.f.
probability = 0.0001

chi-squared with ties = 91.216 with 1 d.f.
probability = 0.0001
```

Le test de Kruskal-Wallis souligne une différence statistiquement significative entre les deux populations de l'échantillon 3.

```
. ranksum nb_claims, by(type_patent)
Two-sample Wilcoxon rank-sum (Mann-Whitney) test
```

type_patent	obs	rank sum	expected
2	2603	6009328.5	6494485
3	2386	6438226.5	5953070
combined	4989	12447555	12447555

```
unadjusted variance 2.583e+09
adjustment for ties -2216058.3
-----
adjusted variance 2.580e+09

Ho: nb_cla~s(type_p~t==2) = nb_cla~s(type_p~t==3)
      z = -9.551
      Prob > |z| = 0.0000
```

Le test de Mann-Whitney confirme cette différence significative entre les 2 moyennes, puisque la  $\text{prob} > |z|$  est inférieure au seuil de significativité de 0,05.

#### Échantillon 4

```
. sktest nb_claims, noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	chi2(2)	joint Prob>chi2
nb_claims	0.000	0.000	2711.98	0.0000



Les données de l'échantillon 4 ne présentent pas une tendance de normalité. Nous effectuons donc un test de déviation standard pour voir si nous pouvons tout de même avoir recours aux tests ANOVA et Bartlett's.

```
. sdtest nb_claims, by(type_patent)
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
2	2562	17.11944	.2840732	14.3787	16.5624 17.67647
3	2386	21.0637	.3325037	16.24171	20.41168 21.71573
combined	4948	19.02142	.2193605	15.43026	18.59138 19.45147

```

ratio = sd(2) / sd(3)
Ho: ratio = 1
Ha: ratio < 1
Pr(F < f) = 0.0000
Ha: ratio != 1
2*Pr(F < f) = 0.0000
Ha: ratio > 1
Pr(F > f) = 1.0000
f = 0.7837
degrees of freedom = 2561, 2385

```

Ce test nous permet de rejeter l'hypothèse d'égalité des variances. Il faut donc procéder avec un test de Kruskal-Wallis.

```
. kwallis nb_claims, by(type_patent)
```

Kruskal-wallis equality-of-populations rank test

type_p~t	Obs	Rank Sum
2	2562	5.82e+06
3	2386	6.43e+06

```

chi-squared = 108.855 with 1 d.f.
probability = 0.0001

chi-squared with ties = 108.947 with 1 d.f.
probability = 0.0001

```

Nous trouvons une différence statistiquement significative entre les moyennes de l'échantillon 4 du nombre de revendications de brevets biotechnologiques et des brevets nanobiotechnologiques.

```
. ranksum nb_claims, by(type_patent)
Two-sample Wilcoxon rank-sum (Mann-Whitney) test
```

type_patent	obs	rank sum	expected
2	2562	5815807.5	6339669
3	2386	6428018.5	5904157
combined	4948	12243826	12243826

```

unadjusted variance    2.521e+09
adjustment for ties    -2131400.8
-----
adjusted variance      2.519e+09

Ho: nb_cla~s(type_p~t==2) = nb_cla~s(type_p~t==3)
      z = -10.438
      Prob > |z| = 0.0000

```

Le test de Mann-Whitney démontre aussi un écart significatif entre les deux moyennes.

## Indicateur de qualité # 2: Nombre de citations (nb\_cit\_5\_yrs).

```
. summarize nb_cit_5_yrs, detail
```

Nb_cit_5_yrs				
Percentiles		Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	16418
25%	0	0	Sum of Wgt.	16418
50%	1		Mean	2.824096
		Largest	Std. Dev.	5.234723
75%	3	75		
90%	7	75	Variance	27.40233
95%	11	78	Skewness	5.406336
99%	25	133	Kurtosis	58.8649

Le tableau ci-dessus résume la distribution du nombre de citations 5 ans suivant l'octroi du brevet. Nous désirons effectuer les analyses en gardant un nombre similaire d'observations pour éviter d'introduire un biais de taille. C'est pourquoi nous conservons les mêmes sous-échantillons employés pour évaluer l'indicateur de qualité #1 (nombre de revendications).

## Nombre de citations après 5 ans : nanotechnologie(1) versus biotechnologie(2)

### Échantillon 1

```
. sktest nb_cit_5_yrs , noadjust
```

```
Skewness/Kurtosis tests for Normality
-----+----- joint -----
Variable | Pr(Skewness) Pr(Kurtosis) chi2(2) Prob>chi2
-----+-----
nb_cit_5_yrs | 0.000 0.000 8077.41 0.0000
```

En effectuant un test Skewness-Kurtosis nous constatons qu'il n'y a pas normalité des données. Parce que nous rejetons la normalité des données, nous ne pouvons procéder avec un test ANOVA. Nous optons pour le test de déviation standard (sdtest).

```
. sdtest nb_cit_5_yrs , by(type_patent)
```

```
Variance ratio test
-----+-----
Group | Obs Mean Std. Err. Std. Dev. [95% Conf. Interval]
-----+-----
1 | 3577 3.962818 .1184824 7.086199 3.730518 4.195118
2 | 3476 2.314154 .066725 3.933949 2.18333 2.444978
combined | 7053 3.150291 .0691943 5.811082 3.014649 3.285932
-----+-----
ratio = sd(1) / sd(2) f = 3.2447
Ho: ratio = 1 degrees of freedom = 3576, 3475
Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1
Pr(F < f) = 1.0000 2*Pr(F > f) = 0.0000 Pr(F > f) = 0.0000
```

Nous rejetons l'égalité des variances, ainsi nous effectuons un test Kruskal-Wallis pour détecter une différence entre les populations.

```
. kwallis nb_cit_5_yrs, by(type_patent)
```

```
Kruskal-Wallis equality-of-populations rank test
```

```
+-----+
| type_p~t | Obs | Rank Sum |
+-----+
| 1 | 3577 | 1.35e+07 |
| 2 | 3476 | 1.13e+07 |
+-----+
```

```
chi-squared = 116.985 with 1 d.f.
probability = 0.0001
```

```
chi-squared with ties = 123.920 with 1 d.f.
probability = 0.0001
```

Le test révèle une différence statistiquement significative entre les deux moyennes.

```
. ranksum nb_cit_5_yrs, by(type_patent)
Two-sample Wilcoxon rank-sum (Mann-Whitney) test
```

type_patent	obs	rank sum	expected
1	3577	13540754	12616079
2	3476	11335177	12259852
combined	7053	24875931	24875931

```
unadjusted variance 7.309e+09
adjustment for ties -4.091e+08
adjusted variance 6.900e+09

Ho: nb~5_yrs(type_p~t==1) = nb~5_yrs(type_p~t==2)
      z = 11.132
      Prob > |z| = 0.0000
```

De plus, le test de Mann-Whitney souligne aussi une différence significative entre les deux moyennes (nanotechnologie et biotechnologie).

## Échantillon 2

```
. sktest nb_cit_5_yrs, noadjust
```

Skewness/Kurtosis tests for Normality

Variable	Pr(Skewness)	Pr(Kurtosis)	joint chi2(2)	Prob>chi2
nb_cit_5_yrs	0.000	0.000	7947.18	0.0000

En procédant avec le test de Skewness-Kurtosis nous obtenons un résultat qui ne nous permet pas de procéder avec un test ANOVA. Nous faisons donc un test de déviation standard.

```
. sdtest nb_cit_5_yrs, by(type_patent)
Variance ratio test
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
1	3577	3.962818	.1184824	7.086199	3.730518 4.195118
2	3496	2.444794	.0734676	4.343918	2.30075 2.588838
combined	7073	3.212498	.0706383	5.940762	3.074026 3.35097

```
ratio = sd(1) / sd(2)          f = 2.6611
Ho: ratio = 1                 degrees of freedom = 3576, 3495

      Ha: ratio < 1           Ha: ratio != 1           Ha: ratio > 1
Pr(F < f) = 1.0000      2*Pr(F > f) = 0.0000      Pr(F > f) = 0.0000
```

Le test de déviation standard ne révèle pas une équivalence des variances. Nous devons donc procéder avec un test de Kruskal-Wallis afin de comparer nos données.

```
. kwallis nb_cit_5_yrs , by( type_patent )
```

Kruskal-Wallis equality-of-populations rank test

type_p~t	Obs	Rank Sum
1	3577	1.36e+07
2	3496	1.15e+07

```
chi-squared = 110.718 with 1 d.f.
probability = 0.0001

chi-squared with ties = 117.372 with 1 d.f.
probability = 0.0001
```

Le test ci-dessus présente des Chi2 avec et sans liens qui sont inférieurs au seuil minimal établi à 0,05. Nous pouvons donc affirmer qu'il y a une différence statistiquement significative entre les moyennes du nombre de citations en nanotechnologie et biotechnologie de cet échantillon.

```
. ranksum nb_cit_5_yrs , by( type_patent )
```

Two-sample wilcoxon rank-sum (Mann-Whitney) test

type_patent	obs	rank sum	expected
1	3577	13555283	12651849
2	3496	11461918	12365352
combined	7073	25017201	25017201

```
unadjusted variance 7.372e+09
adjustment for ties -4.179e+08
adjusted variance 6.954e+09

Ho: nb~5_yrs(type_p~t==1) = nb~5_yrs(type_p~t==2)
      z = 10.834
      Prob > |z| = 0.0000
```

Le test de Mann-Whitney confirme la différence entre les deux moyennes.

### Échantillon 3

```
. sktest nb_cit_5_yrs , noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	chi2(2)	joint Prob>chi2
nb_cit_5_yrs	0.000	0.000	8024.44	0.0000

Avec le test de Skewness-Kurtosis présenté ci-dessus, nous ne pouvons affirmer qu'il y a normalité des données. Nous procédons alors avec un test de déviation standard au lieu d'un test ANOVA.

```
. sdtest nb_cit_5_yrs , by( type_patent )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
1	3577	3.962818	.1184824	7.086199	3.730518 4.195118
2	3465	2.357287	.0670928	3.949365	2.225742 2.488833
combined	7042	3.17282	.0693018	5.815571	3.036968 3.308673

```

      ratio = sd(1) / sd(2)                                f = 3.2194
Ho: ratio = 1                                degrees of freedom = 3576, 3464

      Ha: ratio < 1                                Ha: ratio != 1                                Ha: ratio > 1
Pr(F < f) = 1.0000                                2*Pr(F > f) = 0.0000                                Pr(F > f) = 0.0000

```

Le test sdtest révèle aussi une non-égalité des variances; nous passons donc au test Kruskal-Wallis.

```
. kwallis nb_cit_5_yrs , by( type_patent )
```

Kruskal-Wallis equality-of-populations rank test

type_p~t	Obs	Rank Sum
1	3577	1.35e+07
2	3465	1.13e+07

```

chi-squared = 105.772 with 1 d.f.
probability = 0.0001

chi-squared with ties = 111.914 with 1 d.f.
probability = 0.0001

```

Tout comme pour les échantillons 1 et 2 précédents, le test de Kruskal-Wallis identifie une différence statistiquement significative entre les deux moyennes comparées.

```
. ranksum nb_cit_5_yrs ,by( type_patent )
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

type_patent	obs	rank sum	expected
1	3577	13473573	12596406
2	3465	11324830	12201998
combined	7042	24798403	24798403

unadjusted variance 7.274e+09  
 adjustment for ties -3.992e+08  
 adjusted variance 6.875e+09

Ho: nb~5\_yrs(type\_p~t==1) = nb~5\_yrs(type\_p~t==2)  
 z = 10.579  
 Prob > |z| = 0.0000

Le test de Mann-Whitney permet de confirmer les résultats obtenus précédemment.

### Nombre de citations après 5 ans : nanotechnologie(1) versus nanobiotechnologie(3)

```
. sktest nb_cit_5_yrs , noadjust
```

Skewness/Kurtosis tests for Normality

Variable	Pr(Skewness)	Pr(Kurtosis)	joint chi2(2)	Prob>chi2
nb_cit_5_yrs	0.000	0.000	6228.00	0.0000

Le test de Skewness-Kurtosis relève une non-normalité des données, nous devons donc avoir recours au test de déviation standard pour déterminer s'il est toujours possible de procéder avec le test ANOVA.

```
. sdtest nb_cit_5_yrs , by( type_patent )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
1	3577	3.962818	.1184824	7.086199	3.730518 4.195118
3	2404	3.091514	.124518	6.105191	2.84734 3.335688
combined	5981	3.612607	.0869212	6.722217	3.44221 3.783003

ratio = sd(1) / sd(3) f = 1.3472  
 Ho: ratio = 1 degrees of freedom = 3576, 2403

Ha: ratio < 1 Pr(F < f) = 1.0000  
 Ha: ratio != 1 2\*Pr(F > f) = 0.0000  
 Ha: ratio > 1 Pr(F > f) = 0.0000

Nous n'acceptons pas l'hypothèse selon laquelle il y a équivalence des variances au seuil de significativité de 5%. Ainsi, nous procédons avec un test de Kruskal-Wallis.

```
. kwallis nb_cit_5_yrs , by( type_patent )
```

Kruskal-Wallis equality-of-populations rank test

type_p~t	Obs	Rank Sum
1	3577	1.12e+07
3	2404	6.67e+06

```
chi-squared = 62.189 with 1 d.f.
probability = 0.0001

chi-squared with ties = 65.696 with 1 d.f.
probability = 0.0001
```

Le nombre de citations après 5 ans de brevets en nanotechnologie n'est pas statistiquement équivalent au nombre de citations de ceux en nanobiotechnologie. Parce que nous comparons deux groupes, il est possible d'avoir recours au test de Mann-Whitney.

```
. ranksum nb_cit_5_yrs , by( type_patent )
```

Two-sample wilcoxon rank-sum (Mann-Whitney) test

type_patent	obs	rank sum	expected
1	3577	11215124	10698807
3	2404	6674047	7190364
combined	5981	17889171	17889171

```
unadjusted variance 4.287e+09
adjustment for ties -2.289e+08
adjusted variance 4.058e+09

Ho: nb~5_yrs(type_p~t==1) = nb~5_yrs(type_p~t==3)
      z = 8.105
Prob > |z| = 0.0000
```

Le test de Mann-Whitney révèle que les moyennes ne sont pas égales, confirmant ainsi les résultats obtenus avec le test de Kruskal-Wallis. Le nombre de citations après une période de 5 ans n'est pas pareil en nanotechnologie et en nanobiotechnologie.

### **Nombre de citations après 5 ans : biotechnologie(2) versus nanobiotechnologie(3)**

Comme pour certains des tests précédents, nous séparons le groupe de données de biotechnologie en 4 échantillons pour les comparer avec la nanobiotechnologie afin de garder un ordre de grandeur similaire pour mitiger le risque associé à un effet de taille potentiel.



## Échantillon 1

```
. sktest nb_cit_5_yrs , noadjust
```

```

              Skewness/Kurtosis tests for Normality
      -----+----- joint -----
Variable | Pr(Skewness)  Pr(Kurtosis)  chi2(2)  Prob>chi2
-----+-----
nb_cit_5_yrs |      0.000      0.000    4840.60    0.0000

```

Nous n'acceptons pas la normalité des données de l'échantillon 1. Nous faisons donc un test de déviation standard.

```
. sdtest nb_cit_5_yrs , by( type_patent )
```

Variance ratio test

```

-----+-----
Group | Obs      Mean      Std. Err.   Std. Dev.   [95% Conf. Interval]
-----+-----
2 | 2606    2.527245    .0927388    4.734225    2.345396    2.709094
3 | 2404    3.091514    .124518    6.105191    2.84734    3.335688
-----+-----
combined | 5010    2.798004    .0768868    5.442151    2.647272    2.948736
-----+-----
ratio = sd(2) / sd(3)                                f = 0.6013
Ho: ratio = 1                                degrees of freedom = 2605, 2403

Ha: ratio < 1                                Ha: ratio != 1                                Ha: ratio > 1
Pr(F < f) = 0.0000                        2*Pr(F < f) = 0.0000                        Pr(F > f) = 1.0000

```

Le test de déviation standard révèle une non-égalité des variances fermant ainsi la porte sur la possibilité de procéder avec un test ANOVA. Nous effectuons donc un test de Kruskal-Wallis.

```
. kwallis nb_cit_5_yrs , by( type_patent )
```

Kruskal-Wallis equality-of-populations rank test

```

+-----+
| type_p~t | Obs | Rank Sum |
+-----+
| 2 | 2606 | 6.48e+06 |
| 3 | 2404 | 6.07e+06 |
+-----+

chi-squared = 0.888 with 1 d.f.
probability = 0.3459

chi-squared with ties = 0.962 with 1 d.f.
probability = 0.3267

```

Le test de Kruskal-Wallis indique une probabilité chi-carré élevée avec et sans lien, ce qui indique une différence statistiquement non-significative entre la moyenne du nombre de citations de l'échantillon 1 de biotechnologie et celle de nanobiotechnologie.

```
. ranksum nb_cit_5_yrs , by( type_patent )
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

type_patent	obs	rank sum	expected
2	2606	6481127.5	6529333
3	2404	6071427.5	6023222
combined	5010	12552555	12552555

unadjusted variance 2.616e+09  
 adjustment for ties -2.002e+08  
 adjusted variance 2.416e+09

Ho: nb~5\_yrs(type\_p~t==2) = nb~5\_yrs(type\_p~t==3)  
 z = -0.981  
 Prob > |z| = 0.3267

Le test de Mann-Whitney révèle aussi une similarité des moyennes non-négligeable.

## Échantillon 2

```
. sktest nb_cit_5_yrs , noadjust
```

Skewness/Kurtosis tests for Normality

Variable	Pr(Skewness)	Pr(Kurtosis)	chi2(2)	joint Prob>chi2
nb_cit_5_yrs	0.000	0.000	4831.25	0.0000

Nous répétons la procédure avec l'échantillon 2. La prob>chi2 étant de 0, il n'y a pas normalité des données. Nous effectuons donc un test de déviation standard.

```
. sdtest nb_cit_5_yrs , by(type_patent)
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
2	2641	2.285498	.0735731	3.78097	2.141231 2.429765
3	2404	3.091514	.124518	6.105191	2.84734 3.335688
combined	5045	2.669574	.0709579	5.04001	2.530466 2.808682

ratio = sd(2) / sd(3) f = 0.3835  
 Ho: ratio = 1 degrees of freedom = 2640, 2403

Ha: ratio < 1      Ha: ratio != 1      Ha: ratio > 1  
 Pr(F < f) = 0.0000      2\*Pr(F < f) = 0.0000      Pr(F > f) = 1.0000

Le test de déviation standard indique une non-égalité des variances, il faut donc avoir recours au test de Kruskal-Wallis.

```
. kwallis nb_cit_5_yrs , by( type_patent )
```

Kruskal-wallis equality-of-populations rank test

type_p~t	Obs	Rank Sum
2	2641	6.60e+06
3	2404	6.12e+06

```
chi-squared =      1.282 with 1 d.f.
probability =      0.2575

chi-squared with ties =      1.386 with 1 d.f.
probability =      0.2390
```

Le test de Kruskal-Wallis indique qu'il y aurait une différence non-significative des moyennes, tout comme pour l'échantillon 1.

```
. ranksum nb_cit_5_yrs , by( type_patent )
```

Two-sample wilcoxon rank-sum (Mann-Whitney) test

type_patent	obs	rank sum	expected
2	2641	6604744.5	6663243
3	2404	6123790.5	6065292
combined	5045	12728535	12728535

```
unadjusted variance  2.670e+09
adjustment for ties  -2.007e+08
adjusted variance    2.469e+09

Ho: nb~5_yrs(type_p~t==2) = nb~5_yrs(type_p~t==3)
      z = -1.177
      Prob > |z| = 0.2391
```

Le test de Mann-Whitney propose la même solution. Il ne semble pas y avoir de différence entre la moyenne du nombre de citations après 5 ans en nanobiotechnologie qui est statistiquement égale à celle de l'échantillon 2 de la biotechnologie.

### Échantillon 3

```
. sktest nb_cit_5_yrs , noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	----- joint ----- chi2(2) Prob>chi2	
nb_cit_5_yrs	0.000	0.000	4806.95	0.0000

Les données de l'échantillon 3 de la biotechnologie et de la nanobiotechnologie ne présentent pas une tendance de normalité. Nous procédons donc avec un test de déviation standard.

```
. sdtest nb_cit_5_yrs , by( type_patent )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
2	2615	2.37782	.0771389	3.944659	2.226561	2.52908
3	2404	3.091514	.124518	6.105191	2.84734	3.335688
combined	5019	2.719665	.0720881	5.107075	2.578341	2.860989

ratio = sd(2) / sd(3) f = 0.4175  
Ho: ratio = 1 degrees of freedom = 2614, 2403

Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1  
Pr(F < f) = 0.0000 2\*Pr(F < f) = 0.0000 Pr(F > f) = 1.0000

Le test de déviation standard révèle qu'il n'y a pas égalité des variances. Le test à utiliser devient donc celui de Kruskal-Wallis.

```
. kwallis nb_cit_5_yrs , by( type_patent )
```

Kruskal-Wallis equality-of-populations rank test

type_p~t	Obs	Rank Sum
2	2615	6.52e+06
3	2404	6.07e+06

chi-squared = 0.568 with 1 d.f.  
probability = 0.4511

chi-squared with ties = 0.615 with 1 d.f.  
probability = 0.4330

Nous voyons que les probabilités de chi-carré sont beaucoup plus grandes que le seuil fixé à 0,05. Il est donc possible pour nous d'affirmer qu'il y a égalité statistique des moyennes de l'échantillon 3 du nombre de citations après 5 ans de la biotechnologie avec celles de la nanobiotechnologie.

```
. ranksum nb_cit_5_yrs , by( type_patent )
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

type_patent	obs	rank sum	expected
2	2615	6524997	6563650
3	2404	6072693	6034040
combined	5019	12597690	12597690

unadjusted variance 2.630e+09  
adjustment for ties -1.998e+08  
adjusted variance 2.430e+09

Ho: nb~5\_yrs(type\_p~t==2) = nb~5\_yrs(type\_p~t==3)  
z = -0.784  
Prob > |z| = 0.4330

Le test de Mann-Whitney nous permet de voir que les moyennes sont statistiquement non-différentes.

#### Échantillon 4

```
. sktest nb_cit_5_yrs, noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	----- joint ----- chi2(2)	----- Prob>chi2
nb_cit_5_yrs	0.000	0.000	4720.39	0.0000

Le test de Skewness-Kurtosis démontre ici une non-normalité des données du nombre de citations après 5 ans de la nanobiotechnologie et de l'échantillon 4 de la biotechnologie. Nous procédons alors avec un test de déviation-standard.

```
. sdtest nb_cit_5_yrs , by( type_patent )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
2	2575	2.298641	.0746052	3.785803	2.152348	2.444933
3	2404	3.091514	.124518	6.105191	2.84734	3.335688
combined	4979	2.681462	.0716498	5.055752	2.540997	2.821927

ratio = sd(2) / sd(3) f = 0.3845  
Ho: ratio = 1 degrees of freedom = 2574, 2403

Ha: ratio < 1      Ha: ratio != 1      Ha: ratio > 1  
Pr(F < f) = 0.0000      2\*Pr(F < f) = 0.0000      Pr(F > f) = 1.0000

Nous pouvons voir avec ce test qu'il n'est pas possible d'accepter l'égalité des variances. Comme pour les 3 échantillons précédents, nous ne pouvons avoir recours au test ANOVA et Bartlett's.

```
. kwallis nb_cit_5_yrs , by( type_patent )
```

Kruskal-wallis equality-of-populations rank test

type_p~t	Obs	Rank Sum
2	2575	6.32e+06
3	2404	6.07e+06

chi-squared = 3.011 with 1 d.f.  
probability = 0.0827

chi-squared with ties = 3.273 with 1 d.f.  
probability = 0.0704

Le test de Kruskal-Wallis identifie un chi-carré qui est proche du seuil limite fixé de 5%, mais qui l'excède. Il est donc probable de croire que pour cet échantillon, la différence entre les 2 moyennes n'est pas significative.

```
. ranksum nb_cit_5_yrs , by( type_patent )
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

type_patent	obs	rank sum	expected
2	2575	6323796	6411750
3	2404	6073914	5985960
combined	4979	12397710	12397710

unadjusted variance 2.569e+09  
 adjustment for ties -2.051e+08  
 adjusted variance 2.364e+09

Ho: nb~5\_yrs(type\_p~t==2) = nb~5\_yrs(type\_p~t==3)  
 z = -1.809  
 Prob > |z| = 0.0704

Les résultats ci-dessus démontrent une similarité des moyennes comparées qui est statistiquement significative. Ceci est vrai tant pour cet échantillon que pour l'ensemble des données de moyennes de citations en ce qui concerne la nanotechnologie et la nanobiotechnologie.

### Indicateur d'application d'invention # 1: degré d'application

Pour ce qui est du degré d'application des connaissances associées aux inventions, nous devons encore une fois séparer les groupes en sous-échantillons afin de supprimer le risque associé à un effet de taille. Nous séparons la biotechnologie en 3 groupes pour la comparer à la nanotechnologie et en 4 groupes pour la comparer à la nanobiotechnologie.

### Degré d'application en nanotechnologie(1) versus biotechnologie(2) :

#### Échantillon 1

```
. summarize deg_application , detail
```

deg_application			
-----			
Percentiles		Smallest	
1%	1	0	
5%	2	0	
10%	2	0	Obs 1451
25%	3	0	Sum of wgt. 1451
50%	3.018182		Mean 3.173996
		Largest	Std. Dev. .7020253
75%	3.888889	4	
90%	4	4	Variance .4928395
95%	4	4	Skewness -.8712125
99%	4	4	Kurtosis 4.286933



variances et la possibilité d'utiliser le test ANOVA. Ce dernier indique une différence statistiquement significative entre le degré d'application en nanotechnologie et la biotechnologie.

## Échantillon 2

```
. sktest deg_application , noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	----- joint ----- chi2(2)	Prob>chi2
deg_applic~n	0.000	0.000	180.73	0.0000

Les données de l'échantillon 2 présentent une non-normalité. Il est donc nécessaire de faire un test de déviation standard.

```
. sdtest deg_application, by (type_patent)
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	861	3.04809	.0231676	.679802	3.002619	3.093562
2	606	3.356374	.0271985	.6695477	3.302959	3.409789
combined	1467	3.175439	.0180728	.6922159	3.139987	3.21089

ratio = sd(1) / sd(2)	f = 1.0309
Ho: ratio = 1	degrees of freedom = 860, 605
Ha: ratio < 1	Pr(F < f) = 0.6555
Ha: ratio != 1	2*Pr(F > f) = 0.6890
Ha: ratio > 1	Pr(F > f) = 0.3445

Le test ci-dessus indique une égalité statistique des variances de degré d'application des brevets impliqués. Nous avons donc recours au test ANOVA.

```
. oneway deg_application type_patent, tabulate
```

type_patent	Summary of deg_application		
	Mean	Std. Dev.	Freq.
1	3.0480903	.67980199	861
2	3.3563742	.66954771	606
Total	3.1754387	.69221591	1467

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	33.802383	1	33.802383	74.06	0.0000
within groups	668.650387	1465	.456416646		
Total	702.45277	1466	.479162872		

Bartlett's test for equal variances: chi2(1) = 0.1637 Prob>chi2 = 0.686

Le test de Bartlett's présente ici un prob>chi2=0,686. Nous pouvons donc nous fier aux résultats du test ANOVA. La prob>chi2 étant largement supérieure à 5%, nous pouvons rejeter



l'hypothèse selon laquelle il n'y a pas de différence entre la moyenne du degré d'application en biotechnologie et nanotechnologie.

### Échantillon 3

```
. sktest deg_application , noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	joint chi2(2)	Prob>chi2
deg_applicati~n	0.000	0.000	166.94	0.0000

Les données de l'échantillon 3 présentent une prob>chi2 qui suggère une non-normalité des données. Il n'est donc pas initialement possible de procéder avec un test ANOVA, il faut effectuer un test de déviation standard pour vérifier si les variances sont équivalentes.

```
. sdtest deg_application , by( type_patent )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	861	3.04809	.0231676	.679802	3.002619	3.093562
2	592	3.374035	.0274329	.6674715	3.320157	3.427913
combined	1453	3.180891	.0181892	.6933384	3.145211	3.216571

ratio = sd(1) / sd(2) f = 1.0373  
Ho: ratio = 1 degrees of freedom = 860, 591  
Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1  
Pr(F < f) = 0.6841 2\*Pr(F > f) = 0.6319 Pr(F > f) = 0.3159

Le *sdtest* révèle qu'il n'est pas possible d'affirmer qu'il y a une différence statistiquement significative des variances en question. Nous pouvons ainsi avoir recours au test ANOVA.

```
. oneway deg_application type_patent, tabulate
```

type_patent	Mean	Std. Dev.	Freq.
1	3.0480903	.67980199	861
2	3.3740349	.66747154	592
Total	3.1808909	.69333835	1453

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	37.2689152	1	37.2689152	81.84	0.0000
Within groups	660.733724	1451	.455364386		
Total	698.00264	1452	.480718071		

Bartlett's test for equal variances: chi2(1) = 0.2340 Prob>chi2 = 0.629

Il est possible d'employer le test ANOVA puisque la statistique de Bartlett qui y est associée indique aussi qu'il y a égalité des variances. La  $\text{prob}>F$  étant nulle, nous obtenons un résultat qui implique une différence statistiquement significative entre la moyenne du degré d'application de l'échantillon 3 de la biotechnologie et celle de la nanotechnologie.

Dans le cas des 3 échantillons de nanotechnologie versus biotechnologie, nous pouvons rejeter l'hypothèse selon laquelle les moyennes des degrés d'application sont équivalentes.

### Degré d'application en nanotechnologie(1) versus nanobiotechnologie(3) :

```
. sktest deg_application , noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	----- joint ----- chi2(2)	Prob>chi2
deg_applic~n	0.000	0.000	279.02	0.0000

Pour ce qui est des données du degré d'application de la nanotechnologie et de la nanobiotechnologie, nous ne pouvons affirmer qu'il y a normalité, étant donné les résultats du test de Skewness-Kurtosis.

```
. sdtest deg_application, by( type_patent )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	861	3.04809	.0231676	.679802	3.002619	3.093562
3	802	3.25215	.0255278	.7229382	3.202041	3.30226
combined	1663	3.146501	.0173642	.7081089	3.112443	3.180558

ratio = sd(1) / sd(3)		f = 0.8842
Ho: ratio = 1		degrees of freedom = 860, 801
Ha: ratio < 1	Ha: ratio != 1	Ha: ratio > 1
Pr(F < f) = 0.0381	2*Pr(F < f) = 0.0762	Pr(F > f) = 0.9619

Le test de déviation standard nous permet de voir qu'il est possible de procéder avec un test ANOVA.

```
. oneway deg_application type_patent, tabulate
```

type_patent	Summary of deg_application		
	Mean	Std. Dev.	Freq.
1	3.0480903	.67980199	861
3	3.2521504	.72293824	802
Total	3.1465006	.70810891	1663

Source	Analysis of Variance			F	Prob > F
	SS	df	MS		
Between groups	17.290262	1	17.290262	35.19	0.0000
Within groups	816.066838	1661	.491310559		
Total	833.3571	1662	.501418231		

Bartlett's test for equal variances:  $\chi^2(1) = 3.1402$  Prob> $\chi^2 = 0.076$

Le test de Bartlett possède une prob> $\chi^2$  de 0.076, qui est petite mais tout de même supérieure à la limite fixée de 5%. Pour cette raison, nous acceptons l'utilisation du test ANOVA. La prob>F est de 0.000, nous posons donc qu'il y a une différence significative entre le nombre de degré d'application de la nanotechnologie et de la nanobiotechnologie.

### Degré d'application en biotechnologie(2) versus nanobiotechnologie(3) :

La comparaison du degré d'application de la biotechnologie et de la nanobiotechnologie se fait aussi à l'aide d'échantillons.

#### Échantillon 1

```
. sktest deg_application, noadjust
```

Variable	Skewness/Kurtosis tests for Normality			
	Pr(Skewness)	Pr(Kurtosis)	joint $\chi^2(2)$	Prob> $\chi^2$
deg_applic~n	0.000	0.000	248.46	0.0000

Il y a non-normalité des données, nous devons donc examiner s'il y a équivalence statistique des variances.

```
. sdtest deg_application, by(type_patent)
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
2	462	3.366554	.0332917	.7155783	3.301132	3.431977
3	802	3.25215	.0255278	.7229382	3.202041	3.30226
combined	1264	3.293966	.0203101	.7220797	3.254121	3.333811

ratio = sd(2) / sd(3) f = 0.9797  
Ho: ratio = 1 degrees of freedom = 461, 801

Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1  
Pr(F < f) = 0.4052 2\*Pr(F < f) = 0.8105 Pr(F > f) = 0.5948

En procédant avec un test de déviation standard, nous voyons qu'il n'est pas possible de confirmer que les variances sont égales. Nous effectuons donc un test ANOVA.

```
. oneway deg_application type_patent, tabulate
```

Summary of deg_application			
type_patent	Mean	Std. Dev.	Freq.
2	3.3665542	.71557826	462
3	3.2521504	.72293824	802
Total	3.2939658	.72207974	1264

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	3.83663175	1	3.83663175	7.40	0.0066
Within groups	654.69049	1262	.518772179		
Total	658.527122	1263	.521399146		

Bartlett's test for equal variances: chi2(1) = 0.0611 Prob>chi2 = 0.805

Le test de Bartlett confirme une différence non-significative des variances et permet l'utilisation du résultat du test ANOVA. La prob>F étant très petite, nous acceptons l'hypothèse selon laquelle il y a une différence statistiquement significative entre les moyennes des deux domaines.

## Échantillon 2

```
. sktest deg_application, noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	joint chi2(2)	Prob>chi2
deg_applic~n	0.000	0.000	265.09	0.0000

Les données de l'échantillon 2 présentent une non-normalité. Nous procédons alors avec un test de déviation standard.

```
. sdtest deg_application, by(type_patent)
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
2	440	3.333089	.0341303	.7159239	3.26601	3.400168
3	802	3.25215	.0255278	.7229382	3.202041	3.30226
combined	1242	3.280824	.0204646	.7212129	3.240675	3.320973

ratio = sd(2) / sd(3) f = 0.9807  
Ho: ratio = 1 degrees of freedom = 439, 801  
Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1  
Pr(F < f) = 0.4115 2\*Pr(F < f) = 0.8230 Pr(F > f) = 0.5885

Ce test livre des résultats suggérant que les variances ne soient pas statistiquement différentes. Nous avons donc recours au test ANOVA pour déterminer si les moyennes des degrés d'applications sont différentes.

```
. oneway deg_application type_patent, tabulate
```

type_patent	Summary of deg_application		
	Mean	Std. Dev.	Freq.
2	3.3330887	.71592386	440
3	3.2521504	.72293824	802
Total	3.2808242	.72121295	1242

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	1.86128575	1	1.86128575	3.59	0.0585
Within groups	643.642524	1240	.519066551		
Total	645.503809	1241	.520148114		

Bartlett's test for equal variances: chi2(1) = 0.0538 Prob>chi2 = 0.817

Le test de Bartlett indique une grande statistique de prob>chi2, il est donc possible pour nous de nous fier au test ANOVA. Ce-dernier révèle très légèrement qu'il y a une différence non-significative entre le degré d'application des brevets en nanobiotechnologie et en nanotechnologie.

### Échantillon 3

```
. sktest deg_application, noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	joint chi2(2)	Prob>chi2
deg_applic~n	0.000	0.000	249.79	0.0000

Le test de Skewness-Kurtosis ci-dessus identifie une non-normalité des données concernées. Nous effectuons donc un test de déviation standard pour vérifier s'il y a égalité des variances.

```
. sdtest deg_application, by(type_patent)
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
2	446	3.416018	.0293913	.6207059	3.358255 3.473781
3	802	3.25215	.0255278	.7229382	3.202041 3.30226
combined	1248	3.310712	.0195988	.6923678	3.272262 3.349162

ratio = sd(2) / sd(3) f = 0.7372  
Ho: ratio = 1 degrees of freedom = 445, 801

Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1  
Pr(F < f) = 0.0002 2\*Pr(F < f) = 0.0003 Pr(F > f) = 0.9998

Le test ci-dessus indique que nous pouvons accepter l'équivalence des variances, nous devons procéder avec la version non-paramétrique du test ANOVA, qu'est le test de Kruskal-Wallis.

```
. kwallis deg_application, by(type_patent)
```

Kruskal-wallis equality-of-populations rank test

type_p~t	Obs	Rank Sum
2	446	301214.00
3	802	478162.00

chi-squared = 13.825 with 1 d.f.  
probability = 0.0002

chi-squared with ties = 14.708 with 1 d.f.  
probability = 0.0001

Le test de Kruskal-Wallis montre des probabilités de statistiques chi-carré quasi-nulles, i.e. la différence entre les deux catégories est statistiquement significative.

```
. ranksum deg_application, by(type_patent)
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

type_patent	obs	rank sum	expected
2	446	301214	278527
3	802	478162	500849
combined	1248	779376	779376

unadjusted variance 37229776  
adjustment for ties -2235057.2  
adjusted variance 34994718

Ho: deg\_ap~n(type\_p~t==2) = deg\_ap~n(type\_p~t==3)  
z = 3.835  
Prob > |z| = 0.0001

Le test de Mann-Whitney ci-dessus confirme que le fait qu'il y a une différence statistiquement significative entre les deux moyennes comparées.

#### Échantillon 4

```
. sktest deg_application, noadjust
```

```

              Skewness/Kurtosis tests for Normality
+-----+-----+-----+-----+
Variable | Pr(Skewness)  Pr(Kurtosis)  chi2(2)  joint Prob>chi2
+-----+-----+-----+-----+
deg_applic~n |      0.000      0.000      235.55      0.0000

```

Les données contenues dans le quatrième échantillon ne donnent pas d'indication qu'elles suivent une tendance normale, nous procédons donc avec un test de déviation standard.

```
. sdtest deg_application, by(type_patent)
```

Variance ratio test

```

+-----+-----+-----+-----+-----+-----+
Group | Obs | Mean | Std. Err. | Std. Dev. | [95% Conf. Interval]
+-----+-----+-----+-----+-----+-----+
2 | 440 | 3.334097 | .0308404 | .6469129 | 3.273484 | 3.39471
3 | 802 | 3.25215 | .0255278 | .7229382 | 3.202041 | 3.30226
+-----+-----+-----+-----+-----+-----+
combined | 1242 | 3.281181 | .0198 | .6977933 | 3.242336 | 3.320027
+-----+-----+-----+-----+-----+-----+
ratio = sd(2) / sd(3)                                f = 0.8007
Ho: ratio = 1                                         degrees of freedom = 439, 801
+-----+-----+-----+-----+-----+-----+
Ha: ratio < 1          Ha: ratio != 1          Ha: ratio > 1
Pr(F < f) = 0.0046      2*Pr(F < f) = 0.0092      Pr(F > f) = 0.9954

```

Le tableau ci-dessus révèle qu'il n'est pas possible d'affirmer que les variances sont statistiquement équivalentes.

```
. kwallis deg_application, by(type_patent)
```

Kruskal-Wallis equality-of-populations rank test

```

+-----+-----+-----+
| type_p~t | Obs | Rank Sum |
+-----+-----+-----+
| 2 | 440 | 283752.00 |
| 3 | 802 | 488151.00 |
+-----+-----+-----+

```

```
chi-squared = 2.898 with 1 d.f.
probability = 0.0887
```

```
chi-squared with ties = 3.062 with 1 d.f.
probability = 0.0802
```

Le test de Kruskal-Wallis donne une probabilité de chi-carré qui est basse, mais qui se situe tout de même au-dessus du seuil de significativité de 5%. Nous pouvons donc affirmer qu'il n'y a pas de différence statistique entre les moyennes pour cet échantillon.

```
. ranksum deg_application, by(type_patent)
Two-sample wilcoxon rank-sum (Mann-Whitney) test|
-----+-----
type_patent |      obs      rank sum      expected
-----+-----
          2 |      440      283752      273460
          3 |      802      488151      498443
-----+-----
    combined |     1242      771903      771903

unadjusted variance      36552487
adjustment for ties      -1954738.5
-----
adjusted variance      34597748

Ho: deg_ap~n(type_p~t==2) = deg_ap~n(type_p~t==3)
      z =      1.750
      Prob > |z| =      0.0802
```

Le test de Mann-Whitney offre la même conclusion.

### **COMPARAISONS DE BOUCLES D'INNOVATION QUÉBÉCOISES (dans et hors-boucle) DE QUALITÉ (NOMBRE DE REVENDICATIONS, NOMBRE DE CITATIONS APRÈS 5 ANS) ET DEGRÉ D'APPLICATION.**

Il est important de mentionner ici que parce que les sous-échantillonnages des analyses de la sous-section précédente n'ont pas mis en lumière des situations problématiques ou contradictoires, nous optons de réaliser la suite de nos analyses statistiques en utilisant toutes les données.

#### **Indicateur de qualité # 1 : Nombre de revendications (nb\_claims)**

```
. summarize nb_claims , detail

              Nb_claims
-----+-----
Percentiles      Smallest
1%                1
5%                1
10%               1
25%               1      Obs      531
                        Sum of Wgt. 531
50%              13      Mean      17.3032
                        Std. Dev. 16.98813
75%              21      Largest
90%              33      109
95%              47      115
99%              96      123
                        Variance  288.5966
                        Skewness  3.099075
                        Kurtosis 17.23562
```



### Nombre de revendications de brevets en nanotechnologie, dans-boucle versus hors-boucle :

Cette analyse n'est pas faisable puisque nous nous trouvons ici dans une situation où le nombre de brevets exclusivement nanotechnologiques issues de firmes hors-boucle est de 0. Il y a des brevets nanobiotechnologiques, par contre.

### Nombre de revendications de brevet en biotechnologie, dans-boucle versus hors-boucle :

```
. sktest nb_claims , noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	----- joint ----- chi2(2)	Prob>chi2
nb_claims	0.000	0.000	385.14	0.0000

Le tableau ci-dessus contient un test de Skewness-Kurtosis nous informant que les données concernant le nombre de revendications de brevets appartenant à des firmes dans-boucle et hors-boucle dans le domaine de la biotechnologie ne suivent pas de tendance normale. Ne pouvant pas d'abord effectuer un test ANOVA, nous procédons avec un test de déviations standard.

```
. sdtest nb_claims if type_patent==2, by ( qc_researcher_loop )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	297	16.19529	1.023935	17.64617	14.18017	18.2104
2	80	18.0125	1.604532	14.35137	14.81876	21.20624
combined	377	16.5809	.8754873	16.99889	14.85944	18.30237

ratio = sd(1) / sd(2)	f =	1.5119
Ho: ratio = 1	degrees of freedom =	296, 79
Ha: ratio < 1	Ha: ratio != 1	Ha: ratio > 1
Pr(F < f) = 0.9853	2*Pr(F > f) = 0.0295	Pr(F > f) = 0.0147

En vue des résultats ci-dessus, il semblerait que nous ne pouvons confirmer une égalité statistique des variances. Il faut donc laisser le test ANOVA de côté et employer un test de Kruskal-Wallis.

```
. kwallis nb_claims if type_patent ==2, by( qc_researcher_loop )
```

Kruskal-Wallis equality-of-populations rank test

qc_researcher_loop	Obs	Rank Sum
1	297	54412.50
2	80	16840.50

```
chi-squared =      3.955 with 1 d.f.
```

```
probability =      0.0467
```

```
chi-squared with ties =      3.961 with 1 d.f.
```

```
probability =      0.0466
```

En regardant le test de Kruskal-Wallis, nous pouvons affirmer qu'il y a une différence statistiquement significative entre les moyennes du nombre de revendications de brevet biotechnologiques hors-boucle et ceux dans-boucle.

```
. ranksum nb_claims if type_patent ==2, by( qc_researcher_loop )
```

Two-sample wilcoxon rank-sum (Mann-Whitney) test

qc_researcher_loop	obs	rank sum	expected
1	297	54412.5	56133
2	80	16840.5	15120
combined	377	71253	71253

```
unadjusted variance      748440.00
```

```
adjustment for ties      -1060.18
```

```
adjusted variance      747379.82
```

```
Ho: nb_claims(qc_researcher_loop==1) = nb_claims(qc_researcher_loop==2)
```

```
z = -1.990
```

```
Prob > |z| = 0.0466
```

Le test de Mann-Whitney que nous effectuons révèle la même chose. Au niveau de 0,05, la différence entre les deux moyennes du nombre de revendications est significative.

### Nombre de revendications de brevet en nanobiotechnologie, dans-boucle versus hors-boucle :

```
. sktest nb_claims if type_patent==3, noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	joint chi2(2)	Prob>chi2
nb_claims	0.000	0.000	125.73	0.0000

Le test de Skweness-Kurtosis nous permet de rejeter la normalité au seuil de 5%. Nous passons donc au test de déviation standard pour comparer les variances.

```
. sdtest nb_claims if type_patent==3, by(qc_researcher_loop)

Variance ratio test
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	115	15.93043	1.201304	12.88255	13.55066	18.31021
2	109	19.46789	1.683784	17.57922	16.13034	22.80544
combined	224	17.65179	1.030007	15.41574	15.62199	19.68158

```

      ratio = sd(1) / sd(2)                                f = 0.5370
Ho: ratio = 1                                           degrees of freedom = 114, 108

      Ha: ratio < 1                                Ha: ratio != 1                                Ha: ratio > 1
Pr(F < f) = 0.0006                    2*Pr(F < f) = 0.0011                    Pr(F > f) = 0.9994
```

Il en est de même lorsque nous effectuons le test de Mann-Whitney; la différence entre les moyennes du nombre de revendications des brevets nanobiotechnologiques dans-boucle et hors-boucle n'est pas statistiquement significative.

### Indicateur de qualité # 2 : Nombre de citations après 5 ans de brevetage (nb\_cit\_5\_yrs)

```
. summarize nb_cit_5_yrs , detail
```

Nb_cit_5_yrs				
-----				
Percentiles	Smallest			
1%	0	0		
5%	0	0		
10%	0	0	Obs	532
25%	0	0	Sum of Wgt.	532
50%	1	Largest		Mean 1.904135
75%	2	17	Std. Dev.	3.196665
90%	5	18	Variance	10.21866
95%	8	20	Skewness	3.640225
99%	15	32	Kurtosis	23.33925

### Nombre de citations après 5 ans de brevetage en nanotechnologie, dans-boucle versus hors-boucle :

N'ayant pas un nombre suffisant de brevets exclusivement nanotechnologiques n'appartenant pas à des boucles, nous ne pouvons effectuer une comparaison aux brevets dans-boucle.

### Nombre de citations après 5 ans de brevetage en biotechnologie, dans-boucle versus hors-boucle :

```
. sktest nb_cit_5_yrs , noadjust
```

Skewness/Kurtosis tests for Normality				
----- joint -----				
Variable	Pr(Skewness)	Pr(Kurtosis)	chi2(2)	Prob>chi2
-----+-----	-----	-----	-----	-----
nb_cit_5_yrs	0.000	0.000	449.22	0.0000

Le test de Skewness-Kurtosis indique qu'il y a une non-normalité des données. Nous faisons donc un test de déviation standard pour voir si une égalité des variances permettrait l'utilisation d'une ANOVA.

```
. sdtest nb_cit_5_yrs if type_patent==2, by( qc_researcher_loop )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	298	2.053691	.2041073	3.523438	1.652011	2.455371
2	80	1.5625	.2726126	2.438321	1.019878	2.105122
combined	378	1.949735	.1711031	3.326625	1.613299	2.286171

ratio = sd(1) / sd(2) f = 2.0881  
 Ho: ratio = 1 degrees of freedom = 297, 79

Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1  
 Pr(F < f) = 0.9999 2\*Pr(F > f) = 0.0002 Pr(F > f) = 0.0001

Avec le résultat du test ci-dessus, nous ne pouvons dire qu'il y a égalité des variances. Le test de Kruskal-Wallis est donc de mise.

```
. kwallis nb_cit_5_yrs if type_patent==2, by( qc_researcher_loop )
```

Kruskal-Wallis equality-of-populations rank test

qc_res~p	Obs	Rank Sum
1	298	57005.50
2	80	14625.50

chi-squared = 0.379 with 1 d.f.  
 probability = 0.5379

chi-squared with ties = 0.414 with 1 d.f.  
 probability = 0.5200

Nous obtenons une grande valeur de probabilité chi-carré, ce qui implique qu'il n'y a pas de différence statistiquement significative entre le nombre de citations après 5 ans pour les brevets hors-boucle et dans-boucle en biotechnologie.

```
. ranksum nb_cit_5_yrs if type_patent==2, by( qc_researcher_loop )
```

Two-sample wilcoxon rank-sum (Mann-Whitney) test

qc_research~p	obs	rank sum	expected
1	298	57005.5	56471
2	80	14625.5	15160
combined	378	71631	71631

unadjusted variance 752946.67  
 adjustment for ties -62739.39

adjusted variance 690207.28

Ho: nb~5\_yrs(qc\_res~p==1) = nb~5\_yrs(qc\_res~p==2)  
 z = 0.643  
 Prob > |z| = 0.5200

Le test de Mann-Whitney souligne aussi le fait que la différence est non-significative entre les deux moyennes.

### Nombre de citations après 5 ans de brevetage en nanobiotechnologie, dans-boucle versus hors-boucle :

```
. sktest nb_cit_5_yrs if type_patent==3, noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	----- joint ----- chi2(2)	Prob>chi2
nb_cit_5_yrs	0.000	0.000	145.17	0.0000

En effectuant le test de Skewness-Kurtosis, nous pouvons constater qu'il n'y a pas normalité des données. Nous devons alors nous fier au test de déviation standard pour voir s'il y a égalité des variances, i.e. s'il est possible d'utiliser le test ANOVA.

```
. sdtest nb_cit_5_yrs if type_patent==3, by( qc_researcher_loop )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	115	1.843478	.2437913	2.614371	1.360529	2.326427
2	109	1.66055	.2859314	2.985211	1.093785	2.227316
combined	224	1.754464	.1868246	2.796134	1.386297	2.122632

ratio = sd(1) / sd(2) f = 0.7670  
Ho: ratio = 1 degrees of freedom = 114, 108  
Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1  
Pr(F < f) = 0.0815 2\*Pr(F < f) = 0.1630 Pr(F > f) = 0.9185

Le sdtest révèle que nous ne pouvons confirmer une différence statistiquement significative des variances. Ceci implique que l'utilisation du test ANOVA est acceptable.

```
. oneway nb_cit_5_yrs qc_researcher_loop if type_patent==3, tabulate
```

Qc_researcher_loop	Summary of Nb_cit_5_yrs		
	Mean	Std. Dev.	Freq.
1	1.8434783	2.6143709	115
2	1.6605505	2.985211	109
Total	1.7544643	2.7961341	224

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	1.87256005	1	1.87256005	0.24	0.6256
Within groups	1741.62298	222	7.84514854		
Total	1743.49554	223	7.81836563		

Bartlett's test for equal variances: chi2(1) = 1.9418 Prob>chi2 = 0.163

Le test de Bartlett étant supérieur au seuil de significativité, nous pouvons accepter le résultat du test ANOVA. Ce-dernier indique qu'il y a une différence non-significative entre le nombre de citations après 5 ans des brevets dans-boucle et hors-boucle.

### Comparaison du degré d'application en nanotechnologie selon la position par rapport à la moyenne:

```
. sktest split_gap_by_mean , noadjust
```

```

              Skewness/Kurtosis tests for Normality
      Variable | Pr(Skewness)  Pr(Kurtosis)      chi2(2)      joint Prob>chi2
-----+-----
split_gap~ean |      0.000          .          .          .

```

Le test de Skewness-Kurtosis révèle une non-normalité des données. Nous passons donc au test de déviation standard.

```
. sdtest deg_application if type_patent ==1, by( split_gap_by_mean )
```

Variance ratio test

```

-----+-----
      Group |      Obs      Mean      Std. Err.      Std. Dev.      [95% Conf. Interval]
-----+-----
          1 |         16  2.922265      .079492      .317968      2.752832      3.091698
          2 |         24  2.829945      .0843649     .4133017      2.655423      3.004467
-----+-----
combined |         40  2.866873      .0595237     .3764611      2.746475      2.987271
-----+-----
      ratio = sd(1) / sd(2)
Ho: ratio = 1
      f = 0.5919
      degrees of freedom = 15, 23

      Ha: ratio < 1      Ha: ratio != 1      Ha: ratio > 1
Pr(F < f) = 0.1482      2*Pr(F < f) = 0.2965      Pr(F > f) = 0.8518

```

Le test ci-dessus ne révèle pas de différence significative entre les variances des données. Nous passons donc au test ANOVA.

```
. oneway deg_application split_gap_by_mean if type_patent==1, tabulate
```

```

split_gap_b |      Summary of deg_application
   y_mean |      Mean      Std. Dev.      Freq.
-----+-----
          1 |  2.9222652     .31796797         16
          2 |  2.829945     .41330168         24
-----+-----
        Total |  2.8668731     .37646112         40

```

```

              Analysis of Variance
      Source      SS      df      MS      F      Prob > F
-----+-----
Between groups    .081821063      1    .081821063    0.57    0.4545
Within groups    5.44537489     38    .143299339
-----+-----
        Total    5.52719595     39    .141722973

```

Bartlett's test for equal variances: chi2(1) = 1.1592 Prob>chi2 = 0.282

Le test ANOVA est fiable puisque la statistique de Bartlett est supérieure au seuil de significativité. Nous constatons qu'il y a une différence non-significative entre les degrés d'application de brevets dont l'intervalle de temps entre financement et brevetage sont supérieurs à la moyenne et ceux inférieurs à la moyenne.

### Comparaison du degré d'application en nanotechnologie selon la position par rapport à la médiane:

```
. sktest deg_application, noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	----- joint ----- chi2(2)	Prob>chi2
deg_applic~n	0.000	0.192	16.44	0.0003

Voyant le tableau ci-dessus, les données ne présentent pas une tendance normale.

```
. sdtest deg_application if type_patent ==1, by( split_gap_by_median )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	13	2.935096	.075891	.273629	2.769743	3.100448
2	27	2.834025	.080411	.4178279	2.668738	2.999312
combined	40	2.866873	.0595237	.3764611	2.746475	2.987271

ratio = sd(1) / sd(2) f = 0.4289  
Ho: ratio = 1 degrees of freedom = 12, 26

Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1  
Pr(F < f) = 0.0630 2\*Pr(F < f) = 0.1260 Pr(F > f) = 0.9370

Parce que les variances ne présentent pas d'écart significatif, nous pouvons effectuer un test ANOVA.

```
. oneway deg_application split_gap_by_median if type_patent==1, tabulate
```

split_gap_b y_median	Summary of deg_application		
	Mean	Std. Dev.	Freq.
1	2.9350957	.27362901	13
2	2.8340252	.41782786	27
Total	2.8668731	.37646112	40

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	.089638794	1	.089638794	0.63	0.4336
Within groups	5.43755716	38	.143093609		
Total	5.52719595	39	.141722973		

Bartlett's test for equal variances: chi2(1) = 2.5212 Prob>chi2 = 0.112



Les résultats du test ANOVA ci-dessus étant fiables, nous confirmons une différence non-significative statistiquement entre le degré d'applications de brevets se trouvant au-dessus de la médiane et ceux se trouvant en dessous de la médiane.

### Comparaison du degré d'application en nanotechnologie selon les quartiles 1 ou 4 (loops):

```
. sktest deg_application, noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	chi2(2)	joint Prob>chi2
deg_applic~n	0.000	0.192	16.44	0.0003

Nous menons un test de Skewness-Kurtosis pour vérifier la normalité des données. Celle-ci n'étant pas validée, nous réalisons un test de déviation standard.

```
. sdtest deg_application if type_patent ==1, by( extremes_gap )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	5	3.032222	.1135347	.2538713	2.716999	3.347445
4	18	2.891423	.1054643	.4474473	2.668913	3.113933
combined	23	2.922032	.0859604	.4122516	2.743761	3.100303

ratio = sd(1) / sd(4) f = 0.3219  
Ho: ratio = 1 degrees of freedom = 4, 17  
Ha: ratio < 1 Pr(F < f) = 0.1406  
Ha: ratio != 1 2\*Pr(F < f) = 0.2812  
Ha: ratio > 1 Pr(F > f) = 0.8594

Le test ci-dessus révèle qu'il n'y a pas de différence statistiquement significative entre les variances en question. Nous effectuons donc un test ANOVA.

```
. oneway deg_application extremes_gap if type_patent==1, tabulate
```

extremes_ga p	Summary of deg_application		
	Mean	Std. Dev.	Freq.
1	3.032222	.25387128	5
4	2.891423	.4474473	18
Total	2.9220316	.41225163	23

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	.077573667	1	.077573667	0.44	0.5120
within groups	3.66135736	21	.17435035		
Total	3.73893103	22	.16995141		

Bartlett's test for equal variances: chi2(1) = 1.4991 Prob>chi2 = 0.221

Le test ANOVA est fiable (c.f. test de Bartlett) et indique qu'il ne semble pas y avoir de différence entre les degrés d'application de boucles dans le premier quartile (ceux ayant un intervalle de temps entre les subventions et l'application de brevet d'une année ou moins) et ceux du quatrième quartile (ceux avec intervalle supérieur ou égale à 6.333) en nanotechnologie.

### Comparaison du degré d'application en biotechnologie selon la position par rapport à la moyenne:

```
. sktest deg_application, noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	----- joint ----- chi2(2)	Prob>chi2
deg_applic~n	0.000	0.192	16.44	0.0003

La probabilité de prob>chi2 étant très petite, nous ne pouvons affirmer qu'il y a normalité des données du degré d'application des brevets en biotechnologie.

```
. sdtest deg_application if type_patent ==2, by( split_gap_by_mean )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	47	3.36394	.1089893	.7471929	3.144555	3.583324
2	30	3.381032	.0989371	.5419008	3.178683	3.583381
combined	77	3.370599	.0764505	.6708504	3.218334	3.522863

ratio = sd(1) / sd(2)	f = 1.9012
Ho: ratio = 1	degrees of freedom = 46, 29
Ha: ratio < 1	Pr(F < f) = 0.9655
Ha: ratio != 1	2*Pr(F > f) = 0.0689
Ha: ratio > 1	Pr(F > f) = 0.0345

Le test déviation standard nous permet d'utiliser le test d'ANOVA car celui-ci indique qu'une égalité des variances est possible.

```
. oneway deg_application split_gap_by_mean if type_patent==2, tabulate
```

split_gap_b y_mean	Summary of deg_application		
	Mean	Std. Dev.	Freq.
1	3.3639395	.74719293	47
2	3.3810317	.54190079	30
Total	3.3705988	.67085042	77

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	.005349635	1	.005349635	0.01	0.9140
Within groups	34.1977123	75	.455969497		
Total	34.2030619	76	.450040288		

Bartlett's test for equal variances:  $\chi^2(1) = 3.3984$  Prob> $\chi^2 = 0.065$

Le test de Bartlett's est à la limite de la significativité, mais nous permet tout de même d'employer le résultat du test ANOVA. Il est évident ici que nous ne pouvons qu'accepter une différence statistiquement non-significative entre les moyennes comparées.

### Comparaison du degré d'application en biotechnologie selon la position par rapport à la médiane:

```
. sktest deg_application
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
deg_applic~n	0.000	0.192	13.86	0.0010

En ce qui concerne les degrés d'application de brevets en biotechnologie et leur position par rapport à la médiane, ils ne suivent pas une tendance de normalité. Ceci étant dit, nous effectuons un test de déviation standard pour justifier l'utilisation d'ANOVA.

```
. sdtest deg_application if type_patent ==2, by( split_gap_by_median )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	34	3.327848	.1343457	.7833634	3.054519	3.601176
2	43	3.404402	.0875439	.5740638	3.227731	3.581073
combined	77	3.370599	.0764505	.6708504	3.218334	3.522863

ratio =  $\text{sd}(1) / \text{sd}(2)$  f = 1.8621  
Ho: ratio = 1 degrees of freedom = 33, 42

Ha: ratio < 1      Ha: ratio != 1      Ha: ratio > 1  
Pr(F < f) = 0.9712      2\*Pr(F > f) = 0.0576      Pr(F > f) = 0.0288

```
. oneway deg_application split_gap_by_median if type_patent==2, tabulate
```

split_gap_by_median	Summary of deg_application	
	Mean	Std. Dev.
1	3.3278478	.78336338
2	3.404402	.57406378
Total	3.3705988	.67085042

Source	Analysis of Variance				
	SS	df	MS	F	Prob > F
Between groups	.111274205	1	.111274205	0.24	0.6222
within groups	34.0917877	75	.454557169		
Total	34.2030619	76	.450040288		

Bartlett's test for equal variances:  $\chi^2(1) = 3.5550$  Prob> $\chi^2 = 0.059$

L'utilisation du test étant justifiée, nous pouvons affirmer que la différence est non significative entre les moyennes de degré d'application avec petit intervalle de temps (entre financement et brevetage) et grand intervalle de temps.

### Comparaison du degré d'application en biotechnologie selon les quartiles 1 ou 4 (loops):

```
. sktest deg_application, noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	joint $\chi^2(2)$	Prob> $\chi^2$
deg_applicat	0.000	0.192	16.44	0.0003

Nous effectuons un test de Skewness-Kurtosis et trouvons qu'il n'y a pas normalité des données.

```
. sdtest deg_application if type_patent ==2, by( extremes_gap )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
1	25	3.335873	.1444669	.7223346	3.037708 3.634038
4	24	3.466567	.1109307	.5434473	3.23709 3.696045
combined	49	3.399887	.0911106	.637774	3.216697 3.583077

ratio =  $sd(1) / sd(4)$  f = 1.7667  
Ho: ratio = 1 degrees of freedom = 24, 23

Ha: ratio < 1 Pr(F < f) = 0.9115      Ha: ratio != 1 2\*Pr(F > f) = 0.1770      Ha: ratio > 1 Pr(F > f) = 0.0885

L'utilisation du test de déviation standard est donc de mise. Ce dernier nous révèle qu'il n'est pas possible de confirmer une différence statistique des variances associées aux degrés de liberté des brevets. Nous pouvons donc avoir recours au test ANOVA pour évaluer s'il y a variation des moyennes.

```
. oneway deg_application extremes_gap if type_patent==2, tabulate
```

extremes_ga		Summary of deg_application		
p		Mean	Std. Dev.	Freq.
1		3.335873	.72233462	25
4		3.4665675	.54344729	24
Total		3.3998866	.63777404	49

Source		SS	df	MS	F	Prob > F
Between groups		.209155549	1	.209155549	0.51	0.4791
Within groups		19.3151193	47	.410959984		
Total		19.5242748	48	.406755725		

Bartlett's test for equal variances:  $\chi^2(1) = 1.8305$  Prob> $\chi^2 = 0.176$

Il n'y a pas de différence entre les degrés d'application de brevets dans le premier quartile et ceux du quatrième quartile en biotechnologie.

### Comparaison du degré d'application en nanobiotechnologie selon la position par rapport à la moyenne:

```
. sktest deg_application, noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	joint $\chi^2(2)$	Prob> $\chi^2$
deg_applic~n	0.000	0.192	16.44	0.0003

Les données de degrés d'application ne suivent pas une tendance de normalité, ainsi nous effectuons un test de déviation standard afin d'examiner leur variance.

```
. sdtest deg_application if type_patent ==3, by( split_gap_by_mean)
```

Variance ratio test|

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	59	3.420663	.0639999	.4915922	3.292553	3.548773
2	20	3.253084	.1246856	.5576111	2.992114	3.514054
combined	79	3.378238	.0574583	.5107007	3.263847	3.492629

ratio =  $sd(1) / sd(2)$  f = 0.7772  
Ho: ratio = 1 degrees of freedom = 58, 19

Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1  
Pr(F < f) = 0.2277 2\*Pr(F < f) = 0.4555 Pr(F > f) = 0.7723

L'égalité des variances, pouvant être acceptée, nous effectuons un test ANOVA pour vérifier l'égalité des moyennes.

```
. oneway deg_application split_gap_by_mean if type_patent==3, tabulate
```

split_gap_b y_mean	Summary of deg_application		
	Mean	Std. Dev.	Freq.
1	3.4206631	.4915922	59
2	3.2530843	.55761108	20
Total	3.3782381	.51070065	79

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	.419462283	1	.419462283	1.62	0.2068
Within groups	19.9241197	77	.258754802		
Total	20.343582	78	.260815154		

Bartlett's test for equal variances:  $\chi^2(1) = 0.4647$  Prob> $\chi^2 = 0.495$

Le résultat du test de Bartlett nous indique que l'utilisation du test ANOVA est appropriée. Nous trouvons qu'il n'y a aucune différence entre les moyennes des degrés d'application de brevets avec un petit intervalle de temps entre financement et brevetage et ceux avec un grand intervalle de temps.

### Comparaison du degré d'application en nanobiotechnologie selon la position par rapport à la médiane:

```
. sktest deg_application, noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	joint $\chi^2(2)$	Prob> $\chi^2$
deg_applic~n	0.000	0.192	16.44	0.0003

La prob> $\chi^2$  étant inférieure au seuil de significativité, il n'est pas possible pour nous de dire qu'il y a normalité des données.

```
. sdtest deg_application if type_patent ==3, by( split_gap_by_median)
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	54	3.424137	.0655416	.4816303	3.292677	3.555596
2	25	3.279098	.1132093	.5660466	3.045445	3.51275
combined	79	3.378238	.0574583	.5107007	3.263847	3.492629
ratio = sd(1) / sd(2)				f = 0.7240		
Ho: ratio = 1				degrees of freedom = 53, 24		
Ha: ratio < 1		Ha: ratio != 1		Ha: ratio > 1		
Pr(F < f) = 0.1628		2*Pr(F < f) = 0.3257		Pr(F > f) = 0.8372		

Le test de déviation standard indique qu'il y a égalité statistique des variances, rendant ainsi l'utilisation du test ANOVA possible.

```
. oneway deg_application split_gap_by_median if type_patent==3, tabulate
```

split_gap_by_ y_median	Summary of deg_application		
	Mean	Std. Dev.	Freq.
1	3.4241365	.4816303	54
2	3.2790975	.5660466	25
Total	3.3782381	.51070065	79

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	.359481355	1	.359481355	1.39	0.2429
Within groups	19.9841007	77	.259533775		
Total	20.343582	78	.260815154		

Bartlett's test for equal variances:  $\chi^2(1) = 0.8803$  Prob> $\chi^2 = 0.348$

Le test ANOVA révèle qu'il y a une différence non significative entre les moyennes du degré d'applications selon la position de l'intervalle de temps par rapport à la médiane.

### Comparaison du degré d'application en nanobiotechnologie selon les quartiles 1 ou 4 (loops):

```
. sktest deg_application, noadjust
```

Skewness/Kurtosis tests for Normality				
Variable	Pr(Skewness)	Pr(Kurtosis)	----- joint ----- chi2(2)	Prob>chi2
deg_applic~n	0.000	0.192	16.44	0.0003

Le sktest indique qu'il n'y a pas normalité des données. Le sdtest est donc de mise pour évaluer l'égalité des variances.

```
. sdtest deg_application if type_patent ==3, by( extremes_gap )
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	29	3.641397	.0689988	.3715702	3.500059	3.782734
4	11	3.118867	.188648	.6256745	2.698533	3.539201
combined	40	3.497701	.0798956	.5053042	3.336097	3.659305

ratio = sd(1) / sd(4) f = 0.3527  
Ho: ratio = 1 degrees of freedom = 28, 10

Ha: ratio < 1      Ha: ratio != 1      Ha: ratio > 1  
Pr(F < f) = 0.0144      2\*Pr(F < f) = 0.0289      Pr(F > f) = 0.9856

Le test de déviation standard soulignant une différence statistiquement significative entre les variances à l'étude, nous ne pouvons employer le test ANOVA et nous nous tournons vers le test de Kruskal-Wallis.

```
. kwallis deg_application if type_patent ==3, by( extremes_gap )
```

Kruskal-Wallis equality-of-populations rank test

extremes_gap	Obs	Rank Sum
1	29	675.50
4	11	144.50

chi-squared = 6.020 with 1 d.f.

probability = 0.0141

chi-squared with ties = 6.146 with 1 d.f.

probability = 0.0132

Le test de Kruskal-Wallis identifie qu'il y a une différence statistiquement significative entre le degré d'application de brevets ayant un petit intervalle de temps entre financement de chercheurs et brevets et ceux en ayant un grand.

```
. ranksum deg_application if type_patent ==3, by( extremes_gap )
```

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

extremes_gap	obs	rank sum	expected
1	29	675.5	594.5
4	11	144.5	225.5
combined	40	820	820

unadjusted variance 1089.92

adjustment for ties -22.39

adjusted variance 1067.53

Ho: deg\_ap~n(extremes\_gap==1) = deg\_ap~n(extremes\_gap==4)

z = 2.479

Prob > |z| = 0.0132

Il y a une différence significative du degré d'application des brevets nanobiotechnologiques du 1<sup>er</sup> quartile versus le 4<sup>ème</sup> quartile. Ceux avec petit intervalle de temps semblent être moins appliqués.



## ANNEXE G – Code XML pour classification de brevets nanotechnologiques (nanostructures et nanomatériaux) en sous-catégories avec la méthode K-NN

```

<? xml version="1.0" encoding="UTF-8" standalone="no"?>

<process version="5.1.004">

  <context>

    <input/>

    <output/>

    <macros/>

  </context>

  <operator activated="true" class="process" compatibility="5.1.004" expanded="true"
name="Process">

    <process expanded="true" height="576" width="882">

      <operator activated="true" class="retrieve" compatibility="5.1.004" expanded="true"
height="60" name="Retrieve" width="90" x="179" y="120">

        <parameter key="repository_entry" value="nano_structures_training_data"/>

      </operator>

      <operator activated="true" class="text:process_document_from_data"
compatibility="5.1.001" expanded="true" height="76" name="Process Documents from Data
(3)" width="90" x="313" y="120">

        <list key="specify_weights"/>

        <process expanded="true" height="648" width="951">

          <operator activated="true" class="text:tokenize" compatibility="5.1.001"
expanded="true" height="60" name="Tokenize (3)" width="90" x="112" y="30"/>

```

```

    <operator activated="true" class="text:stem_porter" compatibility="5.1.001"
expanded="true" height="60" name="Stem (Porter)" width="90" x="380" y="30"/>

    <operator activated="true" class="text:filter_stopwords_english"
compatibility="5.1.001" expanded="true" height="60" name="Filter Stopwords (3)" width="90"
x="648" y="30"/>

    <connect from_port="document" to_op="Tokenize (3)" to_port="document"/>

    <connect from_op="Tokenize (3)" from_port="document" to_op="Stem (Porter)"
to_port="document"/>

    <connect from_op="Stem (Porter)" from_port="document" to_op="Filter
Stopwords (3)" to_port="document"/>

    <connect from_op="Filter Stopwords (3)" from_port="document"
to_port="document 1"/>

    <portSpacing port="source_document" spacing="0"/>

    <portSpacing port="sink_document 1" spacing="0"/>

    <portSpacing port="sink_document 2" spacing="0"/>

</process>

</operator>

<operator activated="true" class="x_validation" compatibility="5.1.004"
expanded="true" height="112" name="Validation" width="90" x="447" y="120">

    <parameter key="number_of_validations" value="5"/>

    <parameter key="sampling_type" value="shuffled sampling"/>

    <process expanded="true" height="576" width="395">

        <operator activated="true" class="k_nn" compatibility="5.1.004" expanded="true"
height="76" name="k-NN" width="90" x="179" y="30">

            <parameter key="k" value="3"/>

        </operator>

```

```

    <connect from_port="training" to_op="k-NN" to_port="training set"/>

    <connect from_op="k-NN" from_port="model" to_port="model"/>

    <portSpacing port="source_training" spacing="0"/>

    <portSpacing port="sink_model" spacing="0"/>

    <portSpacing port="sink_through 1" spacing="0"/>

  </process>

  <process expanded="true" height="576" width="395">

    <operator activated="true" class="apply_model" compatibility="5.1.004"
    expanded="true" height="76" name="Apply Model (2)" width="90" x="112" y="30">

      <list key="application_parameters"/>

    </operator>

    <operator activated="true" class="performance" compatibility="5.1.004"
    expanded="true" height="76" name="Performance" width="90" x="246" y="210"/>

    <connect from_port="model" to_op="Apply Model (2)" to_port="model"/>

    <connect from_port="test set" to_op="Apply Model (2)" to_port="unlabelled
    data"/>

    <connect from_op="Apply Model (2)" from_port="labelled data"
    to_op="Performance" to_port="labelled data"/>

    <connect from_op="Performance" from_port="performance" to_port="averagable
    1"/>

    <portSpacing port="source_model" spacing="0"/>

    <portSpacing port="source_test set" spacing="0"/>

    <portSpacing port="source_through 1" spacing="0"/>

    <portSpacing port="sink_averagable 1" spacing="0"/>

    <portSpacing port="sink_averagable 2" spacing="0"/>

```

```

</process>

</operator>

<operator activated="true" class="retrieve" compatibility="5.1.004" expanded="true"
height="60" name="Retrieve (2)" width="90" x="179" y="300">

  <parameter key="repository_entry" value="nano_training_data"/>

</operator>

<operator activated="true" class="text:process_document_from_data"
compatibility="5.1.001" expanded="true" height="76" name="Process Documents from Data
(2)" width="90" x="313" y="300">

  <list key="specify_weights"/>

  <process expanded="true" height="648" width="951">

    <operator activated="true" class="text:tokenize" compatibility="5.1.001"
expanded="true" height="60" name="Tokenize (2)" width="90" x="45" y="30"/>

    <operator activated="true" class="text:stem_porter" compatibility="5.1.001"
expanded="true" height="60" name="Stem (2)" width="90" x="271" y="30"/>

    <operator activated="true" class="text:filter_stopwords_english"
compatibility="5.1.001" expanded="true" height="60" name="Filter Stopwords (2)" width="90"
x="498" y="30"/>

    <connect from_port="document" to_op="Tokenize (2)" to_port="document"/>

    <connect from_op="Tokenize (2)" from_port="document" to_op="Stem (2)"
to_port="document"/>

    <connect from_op="Stem (2)" from_port="document" to_op="Filter Stopwords
(2)" to_port="document"/>

    <connect from_op="Filter Stopwords (2)" from_port="document"
to_port="document 1"/>

    <portSpacing port="source_document" spacing="0"/>

    <portSpacing port="sink_document 1" spacing="0"/>

```

```

    <portSpacing port="sink_document 2" spacing="0"/>

  </process>

</operator>

<operator activated="true" class="apply_model" compatibility="5.1.004"
expanded="true" height="76" name="Apply Model" width="90" x="581" y="300">

  <list key="application_parameters"/>

</operator>

<connect from_op="Retrieve" from_port="output" to_op="Process Documents from
Data (3)" to_port="example set"/>

<connect from_op="Process Documents from Data (3)" from_port="example set"
to_op="Validation" to_port="training"/>

<connect from_op="Validation" from_port="model" to_op="Apply Model"
to_port="model"/>

<connect from_op="Retrieve (2)" from_port="output" to_op="Process Documents
from Data (2)" to_port="example set"/>

<connect from_op="Process Documents from Data (2)" from_port="example set"
to_op="Apply Model" to_port="unlabelled data"/>

<connect from_op="Apply Model" from_port="labelled data" to_port="result 1"/>

<portSpacing port="source_input 1" spacing="0"/>

<portSpacing port="sink_result 1" spacing="0"/>

<portSpacing port="sink_result 2" spacing="0"/>

</process>

</operator>

</process>

```

## ANNEXE H – Code XML pour évaluer la performance des systèmes d'apprentissage, cas du Support Vector Machine pour la biotechnologie

```

<?xml version="1.0" encoding="UTF-8" standalone="no"?>

<process version="5.1.009">

  <context>

    <input/>

    <output/>

    <macros/>

  </context>

  <operator activated="true" class="process" compatibility="5.1.009" expanded="true"
name="Process">

    <process expanded="true" height="444" width="708">

      <operator activated="true" class="retrieve" compatibility="5.1.009" expanded="true"
height="60" name="Retrieve" width="90" x="56" y="67">

        <parameter key="repository_entry" value="../Bio_classification/bio_training2"/>

      </operator>

      <operator activated="true" class="text:process_document_from_data"
compatibility="5.1.002" expanded="true" height="76" name="Process Documents from Data
(2)" width="90" x="246" y="75">

        <parameter key="keep_text" value="true"/>

        <parameter key="prune_below_absolute" value="2"/>

        <parameter key="prune_above_absolute" value="25"/>

        <list key="specify_weights"/>

      </operator>

    </process>

  </operator>

</process>

```

```

    <operator activated="true" class="text:transform_cases" compatibility="5.1.002"
expanded="true" height="60" name="Transform Cases" width="90" x="45" y="30"/>

    <operator activated="true" class="text:tokenize" compatibility="5.1.002"
expanded="true" height="60" name="Tokenize" width="90" x="180" y="30"/>

    <operator activated="true" class="text:filter_stopwords_english"
compatibility="5.1.002" expanded="true" height="60" name="Filter Stopwords (English)"
width="90" x="313" y="30"/>

    <operator activated="true" class="text:stem_porter" compatibility="5.1.002"
expanded="true" height="60" name="Stem (Porter)" width="90" x="447" y="30"/>

    <operator activated="true" class="text:filter_by_length" compatibility="5.1.002"
expanded="true" height="60" name="Filter Tokens (by Length)" width="90" x="579" y="30">

        <parameter key="min_chars" value="2"/>

        <parameter key="max_chars" value="99"/>

    </operator>

    <connect from_port="document" to_op="Transform Cases" to_port="document"/>

    <connect from_op="Transform Cases" from_port="document" to_op="Tokenize"
to_port="document"/>

    <connect from_op="Tokenize" from_port="document" to_op="Filter Stopwords
(English)" to_port="document"/>

    <connect from_op="Filter Stopwords (English)" from_port="document"
to_op="Stem (Porter)" to_port="document"/>

    <connect from_op="Stem (Porter)" from_port="document" to_op="Filter Tokens
(by Length)" to_port="document"/>

    <connect from_op="Filter Tokens (by Length)" from_port="document"
to_port="document 1"/>

    <portSpacing port="source_document" spacing="0"/>

    <portSpacing port="sink_document 1" spacing="0"/>

```

```

    <portSpacing port="sink_document 2" spacing="0"/>

    </process>

    </operator>

    <operator activated="true" class="optimize_parameters_grid" compatibility="5.1.009"
expanded="true" height="130" name="Optimize Parameters (Grid)" width="90" x="447"
y="75">

        <list key="parameters">

            <parameter key="SVM (2).kernel_type"
value="dot,radial,polynomial,anova,epachnenikov,multiquadric"/>

        </list>

        <process expanded="true" height="444" width="708">

            <operator activated="true" class="x_validation" compatibility="5.1.009"
expanded="true" height="112" name="Validation" width="90" x="246" y="75">

                <parameter key="sampling_type" value="shuffled sampling"/>

                <process expanded="true" height="444" width="329">

                    <operator activated="true" class="classification_by_regression"
compatibility="5.1.009" expanded="true" height="76" name="Classification by Regression"
width="90" x="119" y="30">

                        <process expanded="true" height="444" width="708">

                            <operator activated="true" class="support_vector_machine"
compatibility="5.1.009" expanded="true" height="112" name="SVM (2)" width="90" x="112"
y="75">55

                                <parameter key="kernel_type" value="multiquadric"/>

```

---

<sup>55</sup> Nous ajoutons un vecteur de classification par régression pour prédire une étiquette nominale en utilisant le SVM comme système d'apprentissage. Nous faisons ceci car le SVM ne supporte pas des données de type polynomiales comme étiquette.



```

        </operator>

        <connect from_port="training set" to_op="SVM (2)" to_port="training set"/>

        <connect from_op="SVM (2)" from_port="model" to_port="model"/>

        <portSpacing port="source_training set" spacing="0"/>

        <portSpacing port="sink_model" spacing="0"/>

    </process>

</operator>

<connect from_port="training" to_op="Classification by Regression" to_port="training set"/>

    <connect from_op="Classification by Regression" from_port="model"
to_port="model"/>

    <portSpacing port="source_training" spacing="0"/>

    <portSpacing port="sink_model" spacing="0"/>

    <portSpacing port="sink_through 1" spacing="0"/>

</process>

<process expanded="true" height="444" width="329">

    <operator activated="true" class="apply_model" compatibility="5.1.009"
expanded="true" height="76" name="Apply Model" width="90" x="72" y="28">

        <list key="application_parameters"/>

    </operator>

    <operator activated="true" class="performance" compatibility="5.1.009"
expanded="true" height="76" name="Performance" width="90" x="131" y="185"/>

    <connect from_port="model" to_op="Apply Model" to_port="model"/>

    <connect from_port="test set" to_op="Apply Model" to_port="unlabelled data"/>

    <connect from_op="Apply Model" from_port="labelled data"
to_op="Performance" to_port="labelled data"/>

```

```

<connect from_op="Performance" from_port="performance" to_port="averagable 1"/>

    <portSpacing port="source_model" spacing="0"/>

    <portSpacing port="source_test set" spacing="0"/>

    <portSpacing port="source_through 1" spacing="0"/>

    <portSpacing port="sink_averagable 1" spacing="0"/>

    <portSpacing port="sink_averagable 2" spacing="0"/>

</process>

</operator>

<operator activated="true" class="log" compatibility="5.1.009" expanded="true"
height="76" name="Log" width="90" x="473" y="150">

    <list key="log">

        <parameter key="SVM" value="operator.SVM.parameter.kernel_type"/>

        <parameter key="performance"
value="operator.Performance.value.performance"/>

    </list>

</operator>

<connect from_port="input 1" to_op="Validation" to_port="training"/>

<connect from_op="Validation" from_port="model" to_port="result 2"/>

<connect from_op="Validation" from_port="training" to_port="result 1"/>
<connect from_op="Validation" from_port="averagable 1" to_op="Log" to_port="through 1"/>

<connect from_op="Log" from_port="through 1" to_port="performance"/>

<portSpacing port="source_input 1" spacing="0"/>

<portSpacing port="source_input 2" spacing="0"/>

<portSpacing port="source_input 3" spacing="0"/>

<portSpacing port="sink_performance" spacing="0"/>

```

```

    <portSpacing port="sink_result 1" spacing="0"/>
    <portSpacing port="sink_result 2" spacing="0"/>
    <portSpacing port="sink_result 3" spacing="0"/>
  </process>

</operator>

<connect from_op="Retrieve" from_port="output" to_op="Process Documents from
Data (2)" to_port="example set"/>

<connect from_op="Process Documents from Data (2)" from_port="example set"
to_op="Optimize Parameters (Grid)" to_port="input 1"/>

<connect from_op="Process Documents from Data (2)" from_port="word list"
to_op="Optimize Parameters (Grid)" to_port="input 2"/>

<connect from_op="Optimize Parameters (Grid)" from_port="performance"
to_port="result 2"/>
<connect from_op="Optimize Parameters (Grid)" from_port="parameter" to_port="result 1"/>
<connect from_op="Optimize Parameters (Grid)" from_port="result 1" to_port="result 3"/>

  <portSpacing port="source_input 1" spacing="0"/>
  <portSpacing port="sink_result 1" spacing="0"/>
  <portSpacing port="sink_result 2" spacing="0"/>
  <portSpacing port="sink_result 3" spacing="0"/>
  <portSpacing port="sink_result 4" spacing="0"/>
</process>

</operator>

</process>56

```

---

<sup>56</sup> Généré à l'aide du logiciel *Rapidminer*.

## ANNEXE I – Performance de la classification de brevets en nanotechnologie selon la méthode (K-NN, Naive Bayes, Support Machine Vector)

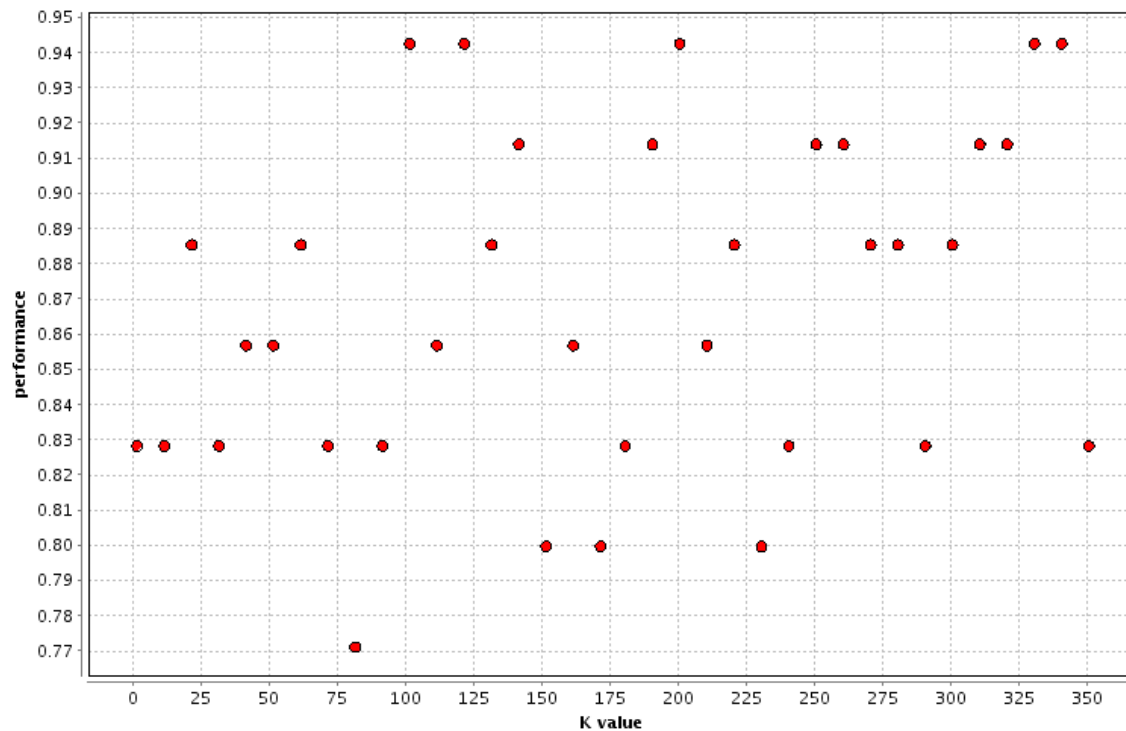


Figure I-1: Optimisation du paramètre k de la classification k-nn des nanoprocessus et nanométrie en nanotechnologie

Tableau I-1: Précision de la classification k-nn des nanoprocessus et nanométrie en nanotechnologie

accuracy: 86.79% +/- 6.55% (mikro: 86.72%)			
	true 0	true 1	class precision
pred. 0	307	47	86.72%
pred. 1	0	0	0.00%
class recall	100.00%	0.00%	

Tableau I-2: Performance par classification k-nn des nanoprocessus et nanométrie en nanotechnologie

K-Value	performance	K-Value	performance	K-Value	performance
1	0.8286	121	0.9429	240	0.8286
11	0.8286	131	0.8857	250	0.9143
21	0.8857	141	0.9143	260	0.9143
31	0.8286	151	0.8000	270	0.8857
41	0.8571	161	0.8571	280	0.8857
51	0.8571	171	0.8000	290	0.8286
61	0.8857	180	0.8286	300	0.8857
71	0.8286	190	0.9143	310	0.9143
81	0.7714	200	0.9429	320	0.9143
91	0.8286	210	0.8571	330	0.9429
101	0.9429	220	0.8857	340	0.9429
111	0.8571	230	0.8000	350	0.8286

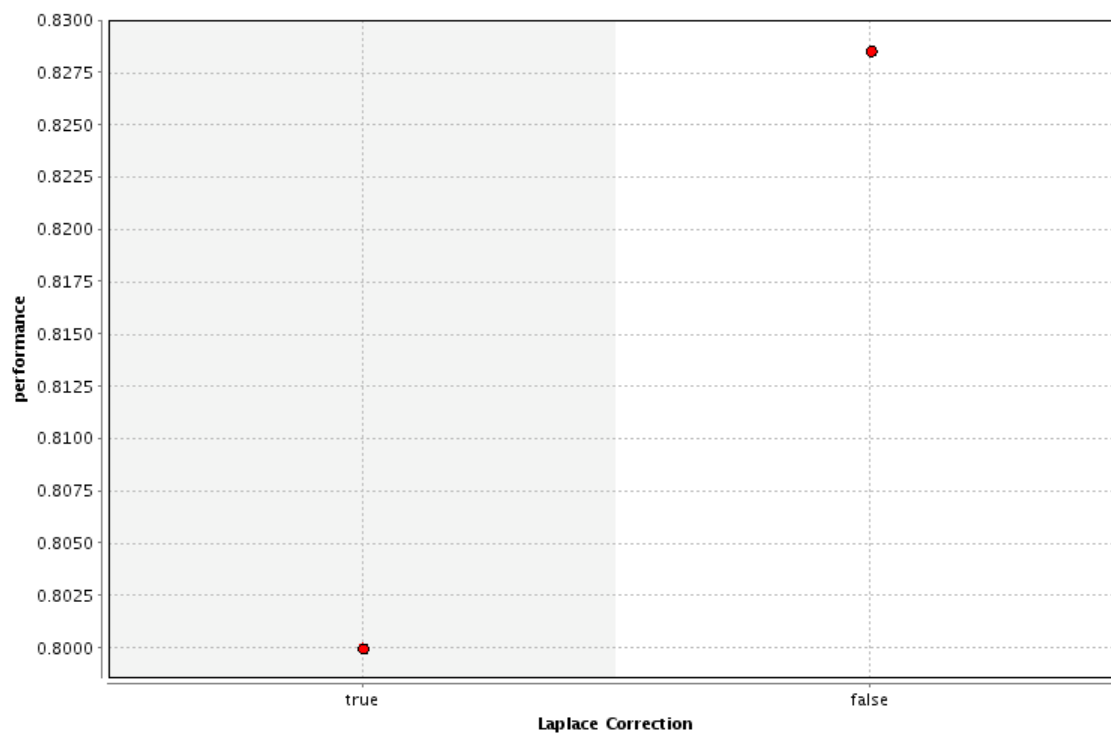


Figure I-2 : Optimisation du paramètre de la correction de Laplace de la classification Naive Bayes des nanoprocessus et nanométrie en nanotechnologie

Tableau I-3: Précision de la classification Naive Bayes des nanoprocessus et nanométrie en nanotechnologie

accuracy: 84.47% +/- 5.71% (mikro: 84.46%)			
	true 0	true 1	class precision
pred. 0	295	43	87.28%
pred. 1	12	4	25.00%
class recall	96.09%	8.51%	

Tableau I-4: Performance par classification Naive Bayes des nanoprocessus et nanométrie en nanotechnologie

Correction de Laplace	performance
TRUE	0.8
FALSE	0.8286

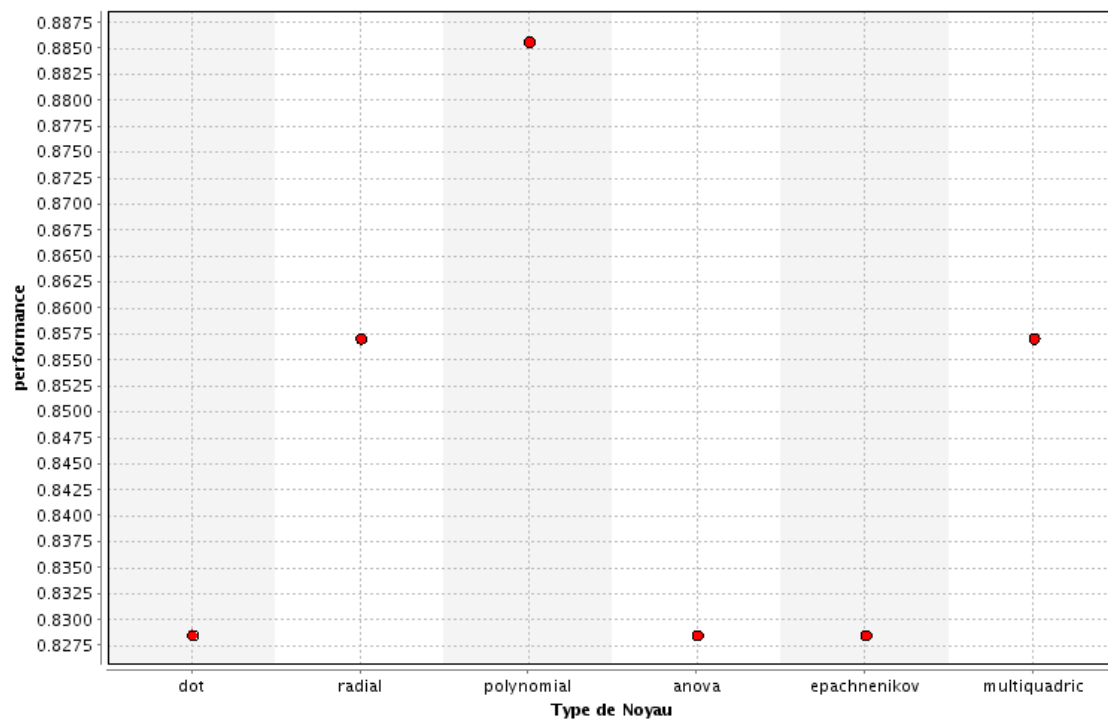


Figure I-3: Optimisation du paramètre de type de noyau de la classification SVM des nanoprocessus et nanométrie en nanotechnologie

Tableau I-5: Précision de la classification SVM des nanoprocessus et nanométrie en nanotechnologie

precision: unknown (positive class: 1)			
	true 0	true 1	class precision
pred. 0	307	47	86.72%
pred. 1	0	0	0.00%
class recall	100.00%	0.00%	

Tableau I-6: Performance par classification SVM des nanoprocessus et nanométrie en nanotechnologie

Type de Noyau	performance
dot	0.8286
radial	0.8571
polynomial	0.8857
anova	0.8286

epachnenikov	0.8286
multiquadric	0.8571

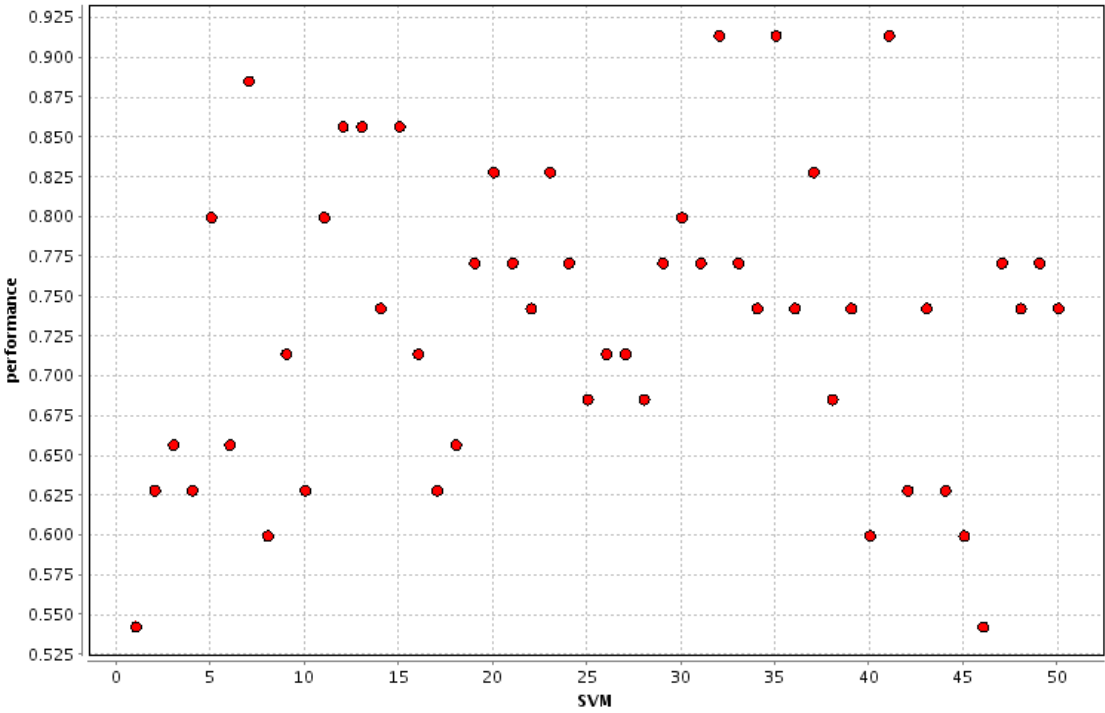


Figure I-4: Optimisation du paramètre k de la classification k-nn de la nanoélectronique et des nanodispositifs en nanotechnologie

Tableau I-7: Précision de la classification k-nn de la nanoélectronique et des nanodispositifs en nanotechnologie

accuracy: 77.09% +/- 6.89% (mikro: 77.12%)			
	true 1	true 0	class precision
pred. 1	169	59	74.12%
pred. 0	22	104	82.54%
class recall	88.48%	63.80%	



Tableau I-8: Performance par classification k-nn de la nanoélectronique et des nanodispositifs en nanotechnologie

k value	performance	k value	performance
1	0.5429	26	0.7143
2	0.6286	27	0.7143
3	0.6571	28	0.6857
4	0.6286	29	0.7714
5	0.8000	30	0.8000
6	0.6571	31	0.7714
7	0.8857	32	0.9143
8	0.6000	33	0.7714
9	0.7143	34	0.7429
10	0.6286	35	0.9143
11	0.8000	36	0.7429
12	0.8571	37	0.8286
13	0.8571	38	0.6857
14	0.7429	39	0.7429
15	0.8571	40	0.6000
16	0.7143	41	0.9143
17	0.6286	42	0.6286
18	0.6571	43	0.7429
19	0.7714	44	0.6286
20	0.8286	45	0.6000
21	0.7714	46	0.5429
22	0.7429	47	0.7714
23	0.8286	48	0.7429
24	0.7714	49	0.7714
25	0.6857	50	0.7429

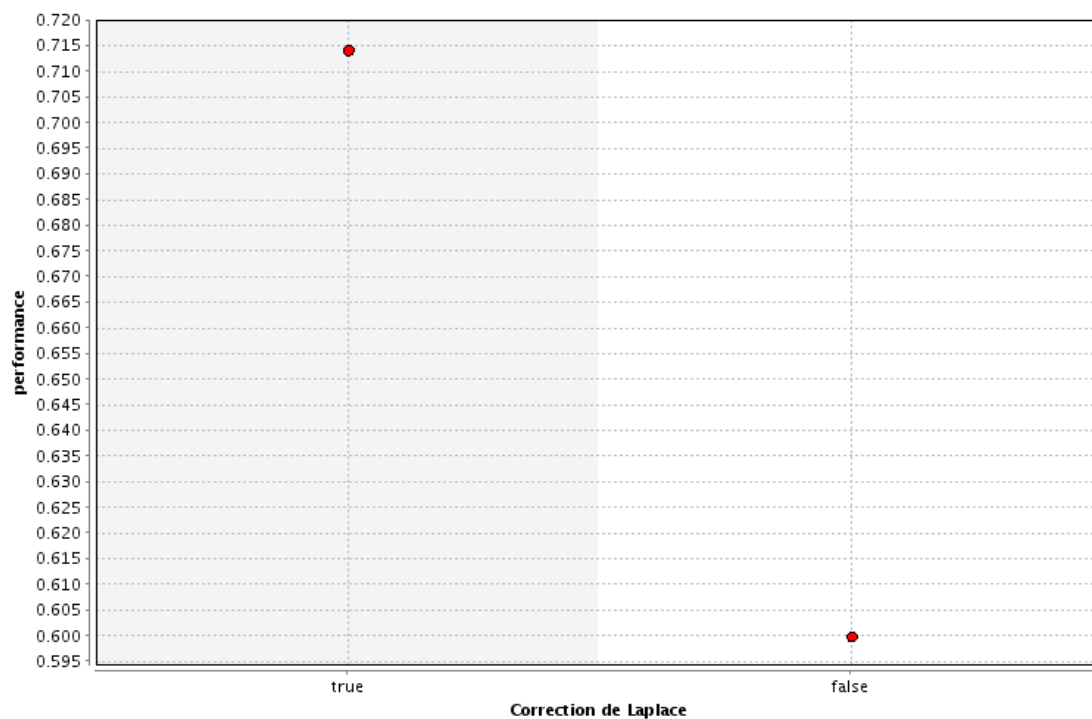


Figure I-5: Optimisation du paramètre de la correction de Laplace de la classification Naive Bayes de la nanoélectronique et des nanodispositifs en nanotechnologie

Tableau I-9: Précision de la classification Naive Bayes de la nanoélectronique et des nanodispositifs en nanotechnologie

accuracy: 70.07% +/- 7.03% (mikro: 70.06%)			
	true 1	true 0	class precision
pred. 1	137	52	72.49%
pred. 0	54	111	67.27%
class recall	71.73%	68.10%	

Tableau I-10: Performance par classification Naive Bayes de la nanoélectronique et des nanodispositifs en nanotechnologie

Correction de Laplace	Performance
TRUE	0.7143
FALSE	0.6

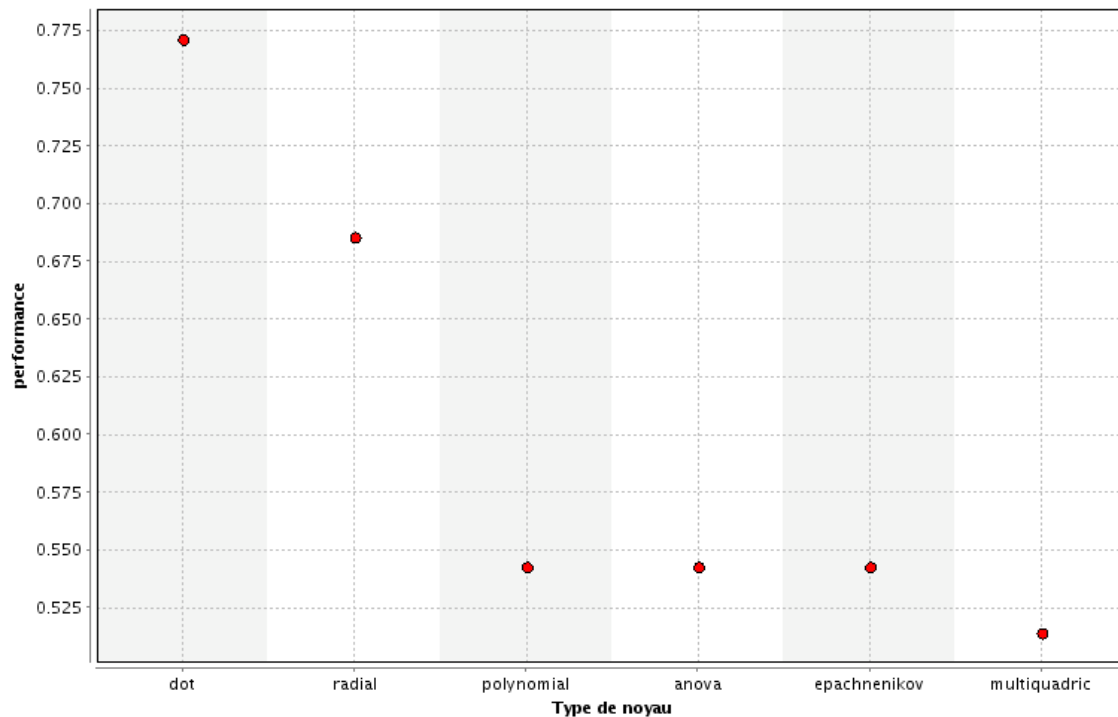


Figure I-6: Optimisation du paramètre de type de noyau de la classification SVM de la nanoélectronique et des nanodispositifs en nanotechnologie

Tableau I-11: Précision de la classification SVM de la nanoélectronique et des nanodispositifs en nanotechnologie

accuracy: 74.33% +/- 4.08% (mikro: 74.29%)			
	true 1	true 0	class precision
pred. 1	159	59	72.94%
pred. 0	32	104	76.47%
class recall	83.25%	63.80%	

Tableau I-12: performance par classification SVM de la nanoélectronique et des nanodispositifs en nanotechnologie

type de Noyau	performance
dot	0.7714

radial	0.6857
polynomial	0.5429
anova	0.5429
epachnenikov	0.5429
multiquadric	0.5143

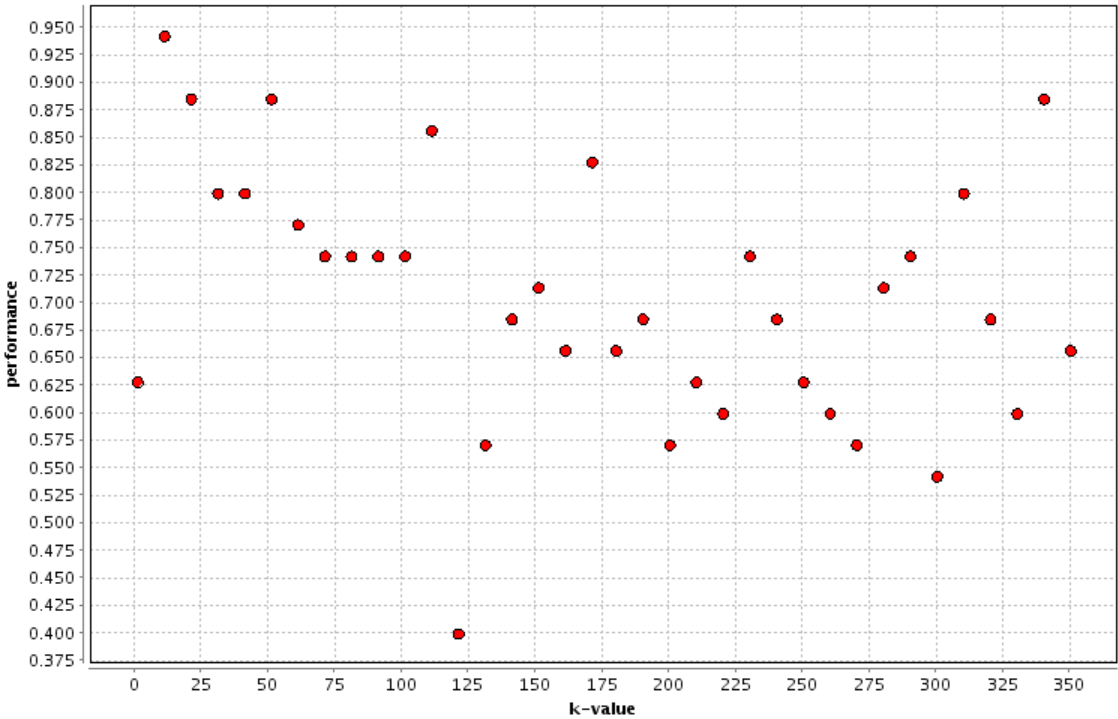


Figure I-7: Optimisation du paramètre k de la classification k-nn des nanomatériaux et des nanostructures en nanotechnologie

Tableau I-13: Précision de la classification k-nn des nanomatériaux et des nanostructures en nanotechnologie

accuracy: 83.30% +/- 5.92% (mikro: 83.33%)			
	true 0	true 1	class precision
pred. 0	226	48	82.48%
pred. 1	11	69	86.25%
class recall	95.36%	58.97%	

Tableau I-14: Performance par classification k-nn des nanomatériaux et des nanostructures en nanotechnologie

K-value	performance	K-value	performance	K-value	performance
1	0.6286	121	0.4000	240	0.6857
11	0.9429	131	0.5714	250	0.6286
21	0.8857	141	0.6857	260	0.6000
31	0.8000	151	0.7143	270	0.5714
41	0.8000	161	0.6571	280	0.7143
51	0.8857	171	0.8286	290	0.7429
61	0.7714	180	0.6571	300	0.5429
71	0.7429	190	0.6857	310	0.8000
81	0.7429	200	0.5714	320	0.6857
91	0.7429	210	0.6286	330	0.6000
101	0.7429	220	0.6000	340	0.8857
111	0.8571	230	0.7429	350	0.6571

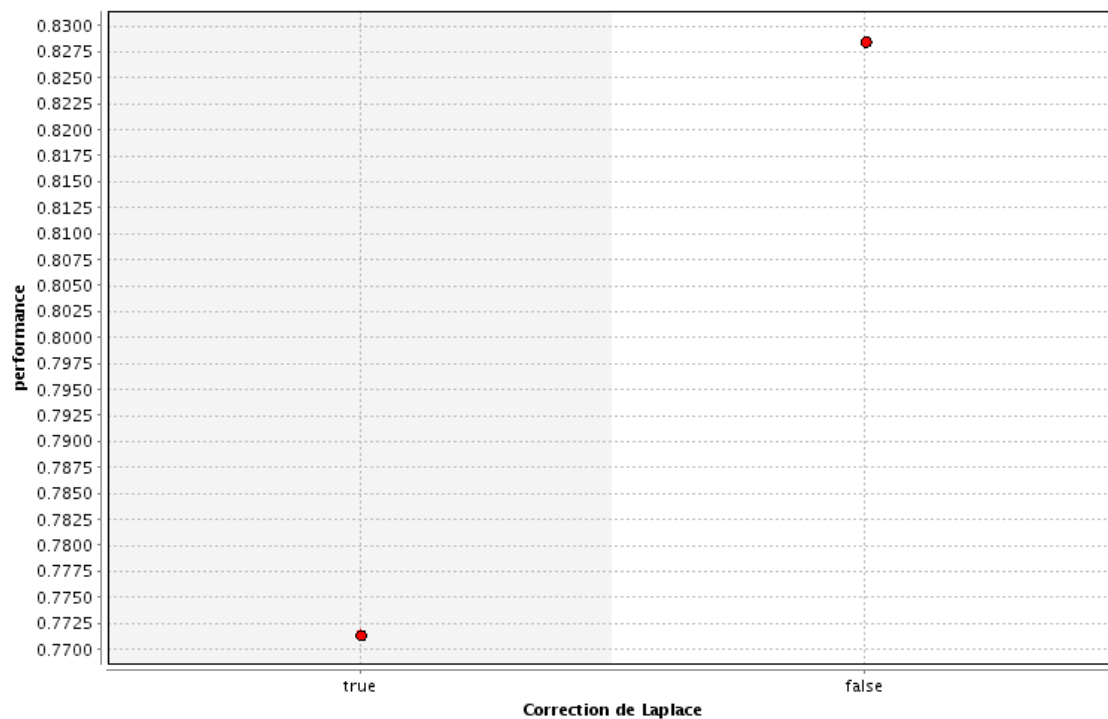


Figure I-8: Optimisation du paramètre de la correction de Laplace de la classification Naive Bayes des nanomatériaux et des nanostructures en nanotechnologie

Tableau I-15: Précision de la classification Naive Bayes des nanomatériaux et des nanostructures en nanotechnologie

accuracy: 80.83% +/- 5.35% (mikro: 80.79%)			
	true 0	true 1	class precision
pred. 0	211	42	83.40%
pred. 1	26	75	74.26%
class recall	89.03%	64.10%	

Tableau I-16: Performance par classification Naive Bayes des nanomatériaux et des nanostructures en nanotechnologie

Correction de Laplace	performanc e
TRUE	0.7714
FALSE	0.8286

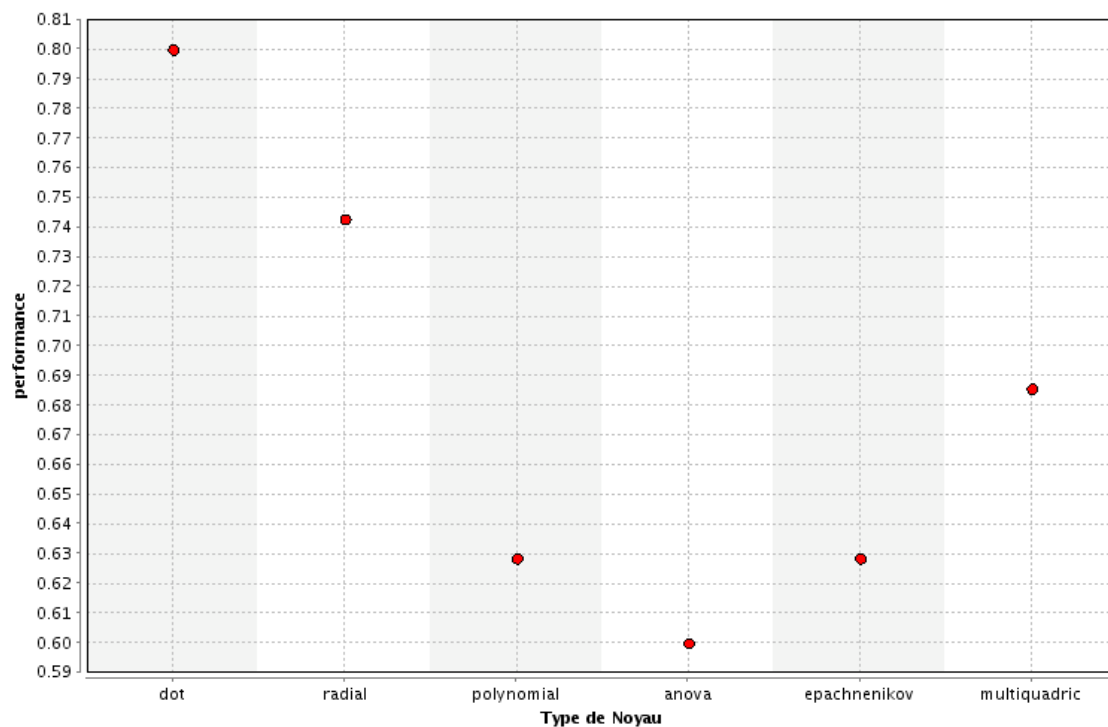


Figure I-9: Optimisation du paramètre k de la classification k-nn des nanomatériaux et des nanostructures en nanotechnologie

Tableau I-17: Précision de la classification SVM des nanomatériaux et des nanostructures en nanotechnologie

accuracy: 81.93% +/- 6.20% (mikro: 81.92%)			
	true 0	true 1	class precision
pred. 0	227	54	80.78%
pred. 1	10	63	86.30%
class recall	95.78%	53.85%	

Tableau I-18: Performance par classification SVM des nanomatériaux et des nanostructures en nanotechnologie

Type de Noyau	performance
dot	0.8000

radial	0.7429
polynomial	0.6286
anova	0.6000
epachnenikov	0.6286
multiquadric	0.6857

## ANNEXE J – Performance de la classification de brevets en biotechnologie selon la méthode (K-NN, Naive Bayes, Support Machine Vector)

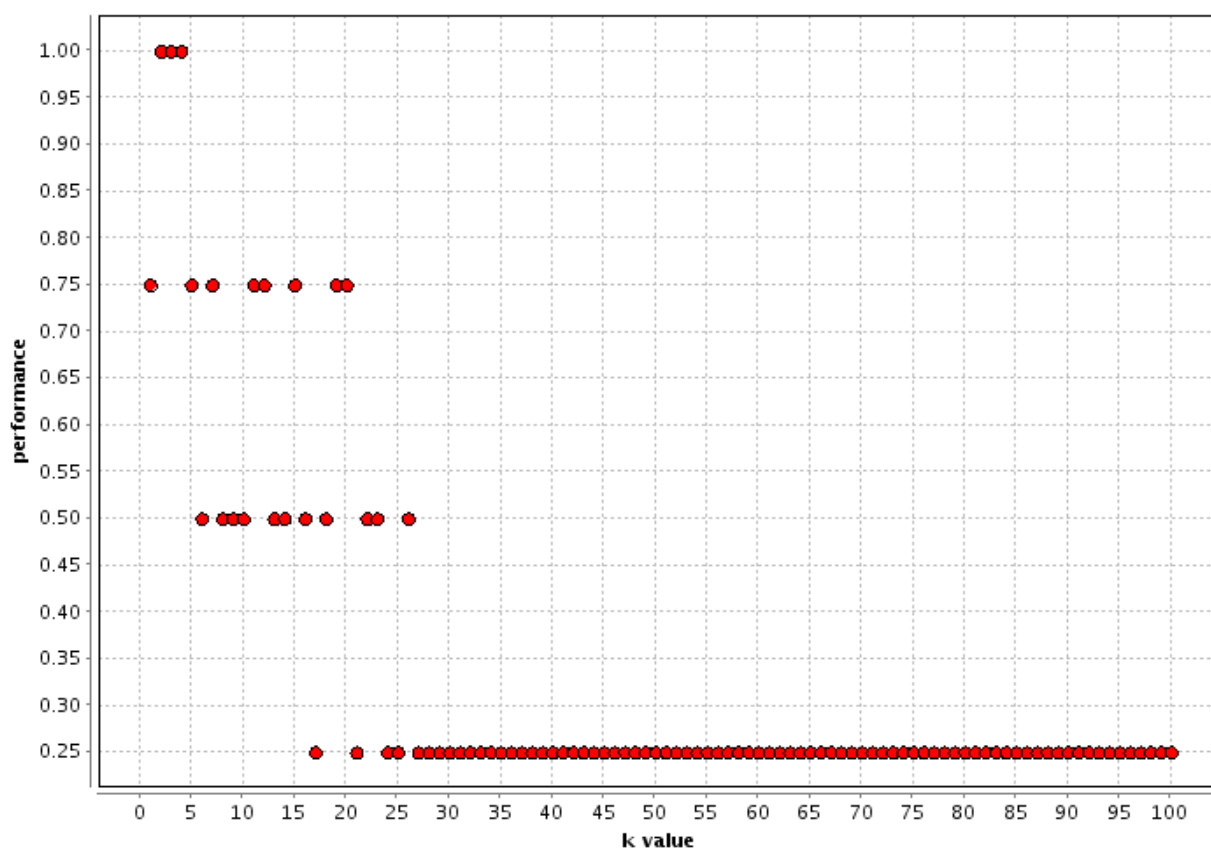


Figure J-1: Optimisation du paramètre k de la classification k-nn de biotechnologie



Tableau J-1: Précision de la classification k-nn en biotechnologie

accuracy: 86.00% +/- 11.58% (mikro: 86.36%)							
	true 1	true 2	true 3	true 4	true 5	true 6	class precision
pred. 1	7	1	0	0	2	1	63.64%
pred. 2	0	6	1	0	0	0	85.71%
pred. 3	0	0	6	0	0	0	100.00%
pred. 4	0	0	0	11	0	0	100.00%
pred. 5	0	0	0	0	6	0	100.00%
pred. 6	1	0	0	0	0	2	66.67%
class recall	87.50%	85.71%	85.71%	100.00%	75.00%	66.67%	

Tableau J-2: Performance par classification k-nn par valeur de k

k-value	performance	k-value	performance	k-value	performance	k-value	performance
1	0.75	26	0.5	51	0.25	76	0.25
2	1	27	0.25	52	0.25	77	0.25
3	1	28	0.25	53	0.25	78	0.25
4	1	29	0.25	54	0.25	79	0.25
5	0.75	30	0.25	55	0.25	80	0.25
6	0.5	31	0.25	56	0.25	81	0.25
7	0.75	32	0.25	57	0.25	82	0.25
8	0.5	33	0.25	58	0.25	83	0.25
9	0.5	34	0.25	59	0.25	84	0.25
10	0.5	35	0.25	60	0.25	85	0.25
11	0.75	36	0.25	61	0.25	86	0.25
12	0.75	37	0.25	62	0.25	87	0.25
13	0.5	38	0.25	63	0.25	88	0.25
14	0.5	39	0.25	64	0.25	89	0.25
15	0.75	40	0.25	65	0.25	90	0.25
16	0.5	41	0.25	66	0.25	91	0.25
17	0.25	42	0.25	67	0.25	92	0.25
18	0.5	43	0.25	68	0.25	93	0.25
19	0.75	44	0.25	69	0.25	94	0.25
20	0.75	45	0.25	70	0.25	95	0.25
21	0.25	46	0.25	71	0.25	96	0.25
22	0.5	47	0.25	72	0.25	97	0.25
23	0.5	48	0.25	73	0.25	98	0.25
24	0.25	49	0.25	74	0.25	99	0.25
25	0.25	50	0.25	75	0.25	100	0.25

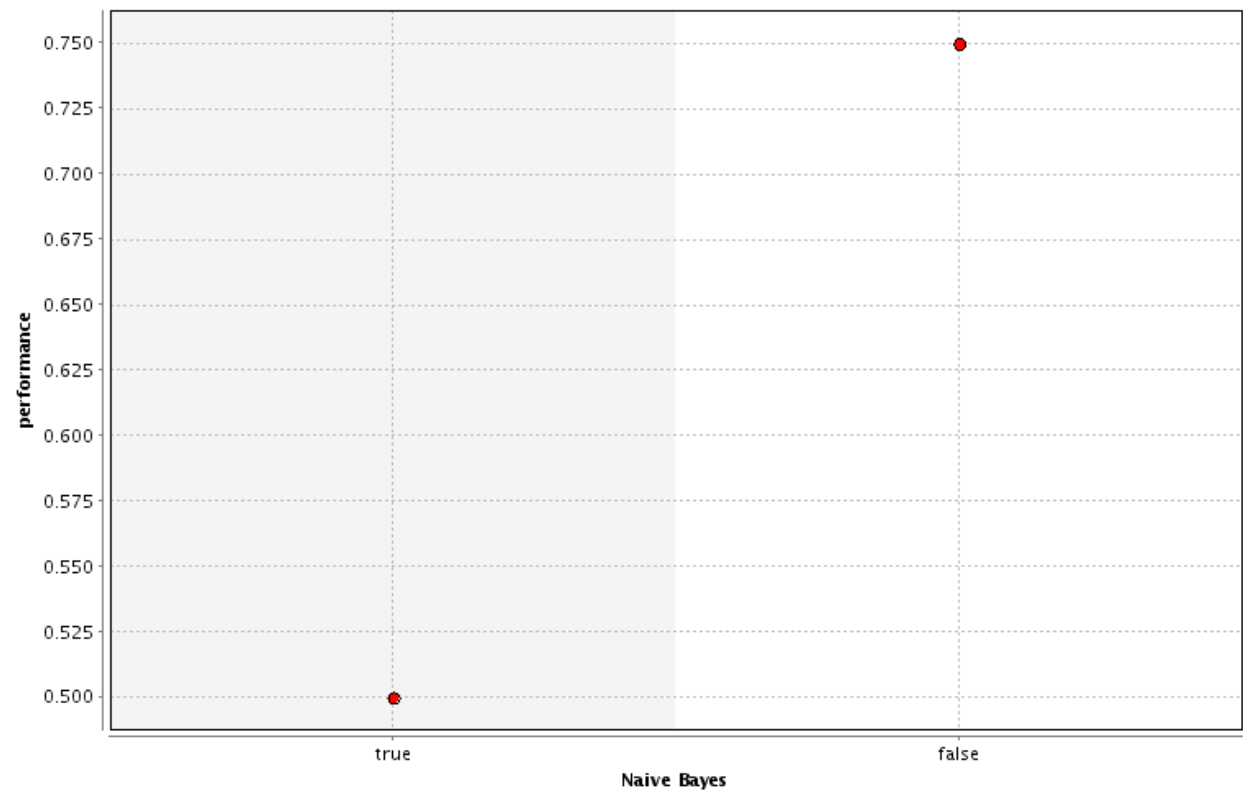


Figure J-2: Optimisation du paramètre de la correction de Laplace(vrai ou faux) de la classification Naive Bayes en biotechnologie

Tableau J-3: Précision de la classification Naïve Bayes en biotechnologie

accuracy: 84.50% +/- 16.35% (mikro: 84.09%)							
	true 1	true 2	true 3	true 4	true 5	true 6	class precision
pred. 1	6	1	0	0	0	2	66.67%
pred. 2	1	5	0	0	1	0	71.43%
pred. 3	0	0	7	0	0	0	100.00%
pred. 4	0	0	0	11	0	0	100.00%
pred. 5	0	0	0	0	7	0	100.00%
pred. 6	1	1	0	0	0	1	33.33%
class recall	75.00%	71.43%	100.00%	100.00%	87.50%	33.33%	

Tableau J-4: Performance par classification Naive Bayes par correction de Laplace en biotechnologie

Correction de Laplace	performance
TRUE	0.5
FALSE	0.75

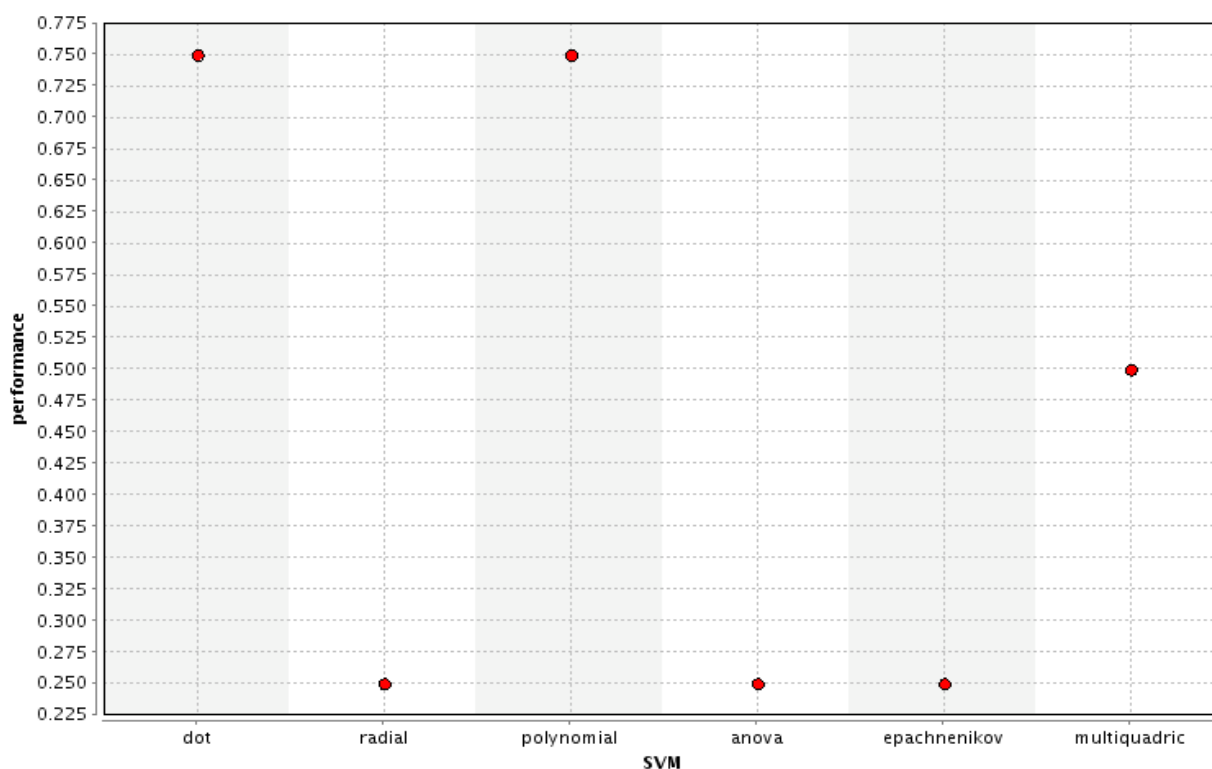


Figure J-3: Optimisation du paramètre de type de noyau de la classification SVM de biotechnologie

Tableau J-5: Précision de la classification SVM en biotechnologie

accuracy: 78.50% +/- 23.56% (mikro: 79.55%)							
	true 1	true 2	true 3	true 4	true 5	true 6	class precision
pred. 1	5	1	0	0	0	1	71.43%
pred. 2	1	6	0	0	2	0	66.67%
pred. 3	0	0	6	0	0	0	100.00%
pred. 4	1	0	0	11	0	1	84.62%
pred. 5	1	0	1	0	6	0	75.00%
pred. 6	0	0	0	0	0	1	100.00%
class recall	62.50%	85.71%	85.71%	100.00%	75.00%	33.33%	

Tableau J-6: Performance par classification SVM par type de noyau en biotechnologie

Kernel type	performance
dot	0.75
radial	0.25
polynomial	0.75
anova	0.25
epachnenikov	0.25
multiquadric	0.5

## ANNEXE K – Options de catégorisation de la nanotechnologie

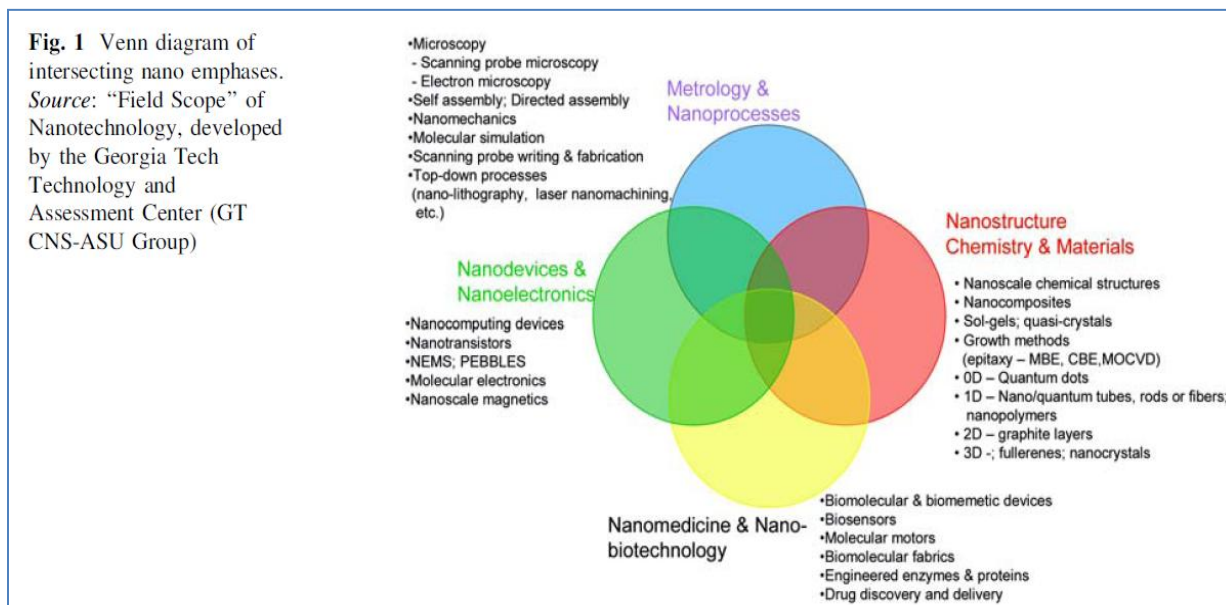


Figure K-1 : Diagramme des sous-sections de la nanotechnologie. Source : Porter et al. (2007). Refining search terms for nanotechnology.

## ANNEXE L – Options de catégorisation de la biotechnologie et applications par certains auteurs

Tableau L-1: Définition de la biotechnologie selon les techniques employées. Source: OECD(2005). A Framework for Biotechnology Statistics. OECD, Paris.

### Box 1. The list-based definition of biotechnology techniques

**DNA/RNA:** Genomics, pharmacogenomics, gene probes, genetic engineering, DNA/RNA sequencing/synthesis/amplification, gene expression profiling, and use of antisense technology.

**Proteins and other molecules:** Sequencing/synthesis/engineering of proteins and peptides (including large molecule hormones); improved delivery methods for large molecule drugs; proteomics, protein isolation and purification, signaling, identification of cell receptors.

**Cell and tissue culture and engineering:** Cell/tissue culture, tissue engineering (including tissue scaffolds and biomedical engineering), cellular fusion, vaccine/immune stimulants, embryo manipulation.

**Process biotechnology techniques:** Fermentation using bioreactors, bioprocessing, bioleaching, biopulping, biobleaching, biodesulphurisation, bioremediation, biofiltration and phytoremediation.

**Gene and RNA vectors:** Gene therapy, viral vectors.

**Bioinformatics:** Construction of databases on genomes, protein sequences; modelling complex biological processes, including systems biology.

**Nanobiotechnology:** Applies the tools and processes of nano/microfabrication to build devices for studying biosystems and applications in drug delivery, diagnostics, etc.

Tableau L-2 : Classification de la biotechnologie selon les applications. Source: OECD(2005). A Framework for Biotechnology Statistics. OECD, Paris.

Table 8. Proposed classification for biotechnology applications		
Broad	Intermediate	Detailed
Human Health	Large molecule therapeutics and monoclonal antibodies (MABs) produced using rDNA technology	-
		Other therapeutics, drug delivery technologies, etc.
	Other therapeutics, artificial substrates, diagnostics and drug delivery technologies, etc.	Substrates (artificial bone, skin etc.)
		Diagnostics
Veterinary health	As above, for veterinary uses	As above
Agriculture	New varieties of genetically modified (GM) plants, animals, and micro-organisms for use in agriculture, aquaculture, and silviculture	GM plants, including fruit trees, flowers, horticultural crops, grains, etc.
		GM animals for agriculture
		GM fish
		GM tree varieties for forestry
		GM micro-organisms for agriculture (including bio pest control)
	New varieties of non-GM plants, animals, and micro-organisms for use in agriculture, aquaculture, silviculture, bio pest control and diagnostics <b>developed using biotechnology techniques</b> (DNA markers, tissue culture, etc.)	Non-GM plants, including fruit trees, flowers, horticultural crops, grains, etc.
		Non-GM animals for agriculture
		Non-GM fish
		Non-GM tree varieties for forestry
		Non-GM micro-organisms for agriculture (including bio pest control)
Natural resources	Applications for mining, petroleum/energy extraction, etc.	Diagnostics
		Mining: extraction using micro-organisms, etc.
		Petroleum/energy: extraction using micro-organisms
Environment	Diagnostics, soil bioremediation, treatment of water, air, and industrial effluents using micro-organisms, clean production processes	Other resource applications
		Diagnosis
		Soil bioremediation, including phytoremediation
		Effluent treatment
Industrial processing	Bioreactors to produce new products (chemicals, food, ethanol, plastics, etc.), biotechnologies to transform inputs (bleaching, biopulping, etc.)	Clean production processes
		Detailed list of specific biotechnologies that are relevant to the firm's sector of activity <sup>3</sup>
Non-specific applications	Research tools, etc.	-
Other		-

Tableau L-3: Catégorisation de brevets selon le code de classification de contenu de la base de données Derwent. Source: Dalpé (2002). Bibliometric Analysis of Biotechnology.

Classification code	Patents			Publications		
	1992-1995	1996-1999	%	1992-1995	1996-1999	%
A Genetic Engineering and Fermentation	13287	20872	57	23739	18022	-24
A1 Nucleic Acid Technology	10512	18224	73	17810	12938	-27
A2 Fermentation	3129	3119	0	6727	6258	-7
B Engineering - Biochemical Engineering	1005	804	-20	3552	2674	-2
C Analysis - Sensors and Analysis	468	441	-6	1682	877	-48
D Pharmaceuticals	9710	15628	61	10235	7991	-22
D1 Antibiotics	678	569	-16	1374	935	-32
D2 Hormones	313	292	-7	448	239	-47
D3 Peptides and Proteins	2413	6674	176	1200	1085	-10
D4 Vaccines	1244	1885	52	1116	799	-28
D5 Other Pharmaceuticals	3152	2401	-24	2797	1609	-42
D6 Antibodies	1384	1918	39	944	684	-28
D7 Clinical Genetic Techniques	1946	7152	268	2977	3099	4
E Agriculture	2083	2959	42	6269	4581	-27
E1 Biological Control Agents	321	295	-8	511	295	-42
E2 Plant Genetic Engineering	661	1571	138	1668	1579	-5
E3 Pesticides	398	432	9	561	695	24
E4 In-Vitro Propagation	589	498	-15	3320	1996	-40
E5 Agriculture, Other	306	443	45	533	300	-44
F Food - Food and Food Additives	1616	1608	-1	2510	2019	-20
G Fuels, Mining and Metal Recovery	564	494	-12	2094	1443	-31
G1 Biofuels and Solvents	406	298	-27	1611	877	-46
G2 Mining and Metal Recovery	162	197	22	489	569	16
H Other Chemicals	1722	1598	-7	3070	2805	-9
H1 Polymers	523	511	-2	823	767	-7
H2 Chiral Compounds	541	530	-2	1244	1167	-6
H3 Miscellaneous Compounds	429	570	33	680	890	31
J Cell Culture	3683	3626	-2	5501	2803	-49
J1 Animal Cell Culture	3445	3472	1	4610	2119	-54
J2 Plant Cell Culture	269	198	-26	921	689	-26
K Biocatalysis	5325	4268	-20	12252	9102	-26
K1 Isolation and Characterization	2032	1653	-19	5567	3869	-31
K2 Application	3656	3095	-15	7276	5886	-19
L Purification – Downstream Processing	930	505	-46	1737	859	-51
M Waste Disposal and the Environment	1507	1868	24	5767	5520	-4
M1 Industrial Waste Disposal	1177	1250	6	3212	2094	-35
M2 Environmental Biotechnology	371	763	106	2720	3689	36
TOTAL	20558	26375	28	41133	29906	-27

Tableau L-4: Catégorisation de brevets selon le code de classification de contenu de la base de données Derwent (Abrégé). Source : Ganerjee, Gupta & Garg (2000). Patent Statistics as indicators of competition, an analysis of patenting in biotechnology. NISTADS, New Delhi.

Table 6 Subject-wise break-up of patents in biotechnology								
Subject code	Subject	Number and percentage of patents						Change in % contribution*
		1982-93		1982-84		1991-93		
		No.	%	No.	%	No.	%	
A1-2	Microbiology	25072	25.62	3238	21.58	8998	30.76	+9.18
B1	Engineering	4467	4.56	187	6.50	753	2.57	-3.93
C1	Chemistry	2325	2.37	335	2.23	520	1.77	-0.46
D1-5	Pharmaceuticals	19437	19.86	2768	18.45	5855	20.21	+1.76
E1-5	Agriculture	4369	4.46	495	3.30	1642	5.61	+2.31
F1	Food Additives and SCP	5969	6.10	1402	9.34	1153	3.94	-5.40
G1	Energy	2180	2.22	645	4.30	346	1.18	-2.12
H1-4	Other Chemicals	3496	3.57	428	2.85	1262	4.31	-1.46
J1-2	Cell Culture	8042	8.21	650	4.23	2867	9.80	+5.57
K1-2	Bio Catalysis	13889	4.19	2123	14.15	4221	14.43	-0.28
L1	Purification	4977	5.09	69	7.79	792	2.70	-5.09
M1	Waste Disposal	3620	3.69	770	5.13	840	2.87	-2.26



Tableau L-5: Secteurs industriels de la biotechnologie. Source : Traoré (2004). Biotechnology Use and Development Survey : methodology, issues and responses. Statistics Canada. Ottawa.

<p><b>Human Health - Bio</b></p> <p>1) <b>Diagnostics</b> (e.g. immunodiagnostics, gene probes, biosensors)</p> <p>2) <b>Therapeutics</b> (e.g. vaccines, immune stimulants, biopharmaceuticals, rational drug design, drug delivery, combinatorial chemistry)</p> <p>3) <b>Gene Therapy</b> (e.g. gene identification, gene constructs, gene delivery)</p> <p><b>Bio-Informatics</b></p> <p>4) <b>Genomics and Molecular Modelling</b> (e.g. DNA/RNA/protein sequencing &amp; databases for humans, plants, animals and microorganisms)</p> <p><b>Ag - Bio</b></p> <p>5) <b>Plant Biotechnology</b> (e.g. tissue culture, embryogenesis, genetic markers, genetic engineering)</p> <p>6) <b>Animal Biotechnology</b> (e.g. diagnostics, therapeutics, embryo transplantation, genetic markers, genetic engineering)</p> <p>7) <b>Biofertilizers/Biopesticides/Bioherbicides/Biological Feed Additives/Microbial pest control</b> (e.g. bacteria, fungi, yeasts)</p> <p>8) <b>Non-Food Applications of Agricultural Products</b> (e.g. fuels, lubricants, commodity and fine chemical feedstocks, cosmetics)</p> <p><b>Food Processing</b></p> <p>9) <b>Bioprocessing</b> (e.g. using enzymes and bacteria culture)</p> <p>10) <b>Functional Foods/Nutraceuticals</b> (e.g. probiotics, unsaturated fatty acids)</p>	<p><b>Aquaculture</b></p> <p>11) <b>Fish health</b> (e.g. diagnostics, therapeutics)</p> <p>12) <b>Broodstock genetics</b> (e.g. tracking superior traits, genetic modification / engineering)</p> <p>13) <b>Bioextraction</b> (e.g. karageenan from seaweed, antifreeze proteins from fish, flavours)</p> <p><b>Mining/Energy/Petroleum/Chemicals</b></p> <p>14) <b>Microbiologically enhanced petroleum/mineral recovery</b></p> <p>15) <b>(Cleaner) Industrial Bioprocessing</b> (e.g. biodesulphurization, bio-cracking, bio-recovery)</p> <p><b>Forest Products</b></p> <p>16) <b>Silviculture</b> (e.g. ectomycorrhizae, tissue culture, somatic embryogenesis, genetic markers, genetic engineering)</p> <p>17) <b>(Cleaner) Industrial Bioprocessing</b> (e.g. biopulping, biobleaching, biological prevention of sapstain)</p> <p><b>Environment</b></p> <p>18) <b>Biofiltration</b> (e.g. treatment of organic emissions to air/water)</p> <p>19) <b>Bioremediation and Phytoremediation</b> (e.g. cleanup of toxic waste sites using microorganisms)</p> <p>20) <b>Diagnostics</b> (e.g. detection of toxic substances using bioindicators, biosensors, immunodiagnostics)</p> <p><b>Other</b></p> <p>21) <b>Custom synthesis- chemical or biological</b> (e.g. peptides, proteins, nucleotides, hormones, growth factors, biochemicals)</p> <p>22) <b>Other</b> (please specify)</p>
--	--

Tableau L-6 : Secteurs d'activités de la Biotechnologie et agrégation. Source : Arundel (1999).  
Diffusion of biotechnologies in Canada: Results from the survey of biotechnology use in Canadian Industries. Statistics Canada.

Table 2.2. Definition of the sector of activity		
Two-digit SIC level	Aggregation to eight sectors	Aggregation to two sectors
Fishing	1. Food	1. Manufacturing
Food		
Beverages		
Tobacco		
Pharmaceuticals	2. Pharmaceuticals	
Non-pharmaceutical chemicals	3. Non pharmaceutical chemicals	
Primary textiles	4. Other	
Textile products		
Leather products		
Fabricated metal products		
Printing and publishing		
Other (includes instruments)		
Mining	5. Mining	2. Resources
Crude petroleum and gas	6. Crude petroleum	
Refined petroleum and coal	7. Refined petroleum	
Forestry services	8. Wood, paper and pulp	
Wood products		
Paper and allied products		

Tableau L-7: Pourcentage d'utilisation des biotechnologies selon catégorisation en secteurs et en applications

<b>Table 4.1. Employee-weighted percentages for the use of 22 biotechnologies in eight sectors<sup>1</sup></b>								
	Resource-based sectors				Manufacturing sectors			
	M	CP	PR	WPP	F	P	NPC	O
Number of respondent firms	49	124	36	201	746	61	234	559
<b>Selection and modification biotechnologies</b>								
Recombinant DNA	-	-	-	-	8	14	1	-
Antibodies/antigens	-	-	-	-	5	15	-	-
Peptide synthesis	-	-	-	-	-	14	-	-
Rational drug design	-	-	-	-	-	13	-	-
Monoclonal antibodies	-	-	-	-	2	15	1	-
Gene probe	-	-	-	-	7	14	2	-
Gene therapy	-	-	-	-	-	-	-	-
DNA amplification	-	-	-	-	7	14	-	-
<b>Environmental biotechnologies</b>								
Bioaugmentation	9	12	48	26	11	-	2	-
Bioremediation	25	39	64	52	10	-	11	2
Bio-reactors	16	-	53	40	9	-	3	-
Phytoremediation	15	14	26	3	-	-	-	-
Biological gas cleaning	-	5	48	-	1	-	2	-
<b>Culture and biological material technologies</b>								
Tissue culture	-	-	-	1	6	15	2	-
Somatic embryo genesis	-	-	-	-	-	-	1	-
Bio-pesticide	-	-	-	7	1	8	1	-
Classical/traditional breeding	3	-	-	3	8	4	1	-
Bioprocessing	-	-	28	1	15	15	-	-
Bio sensing	3	-	-	1	11	28	1	-
Bio-bleaching	-	-	-	1	-	-	-	-
Bio-leaching	14	-	-	-	-	-	-	2
Microbio-inoculants	-	2	-	1	1	-	1	-

<sup>1</sup> Excludes 12 firms with ten or more establishments. M: mining, CP: crude petroleum, PR: petroleum refining, WPP: wood, paper, and pulp, F: food, beverages and tobacco, P: pharmaceuticals, NPC: non-pharmaceutical chemicals, O: other manufacturing.

Light-shaded squares: > 10% - 24.9%, darker shaded squares: > 25%.

## ANNEXE M – Tables de corrélation

Tableau M-1: table de corrélation entre des variables pour les blocs A, B et C

	nbclaims	nbpat	nbcitp5	nbartmoyan3	ltotcontmoyan3	sq_ltotcontmoyan3	ltotsubvmoyan3	sq_ltotsubvmoyan3	lmcliquind23	sq_lmcliquind23	deg_app	nbloop	propuninv	excl_nano_pat	excl_bio_pat	ltotcontmoyan3wu	ltotsubvmoyan3wu	nbartmoyan3wu	codechair33	age	sq_age
nbclaims	1,00																				
nbpat	0,72	1,00																			
nbcitp5	0,48	0,50	1,00																		
nbartmoyan3	0,01	0,02	0,01	1,00																	
ltotcontmoyan3	0,02	-0,08	-0,01	0,10	1,00																
sq_ltotcontmoyan3	0,05	-0,06	-0,01	0,12	0,98	1,00															
ltotsubvmoyan3	-0,03	-0,05	-0,08	0,25	0,28	0,30	1,00														
sq_ltotsubvmoyan3	-0,03	-0,05	-0,08	0,26	0,29	0,31	0,99	1,00													
lmcliquind23	0,03	0,12	0,10	0,22	0,01	0,02	-0,09	-0,08	1,00												
sq_lmcliquind23	0,05	0,14	0,11	0,17	0,01	0,02	-0,09	-0,08	0,95	1,00											
deg_app	0,02	0,04	0,01	0,05	-0,14	-0,14	0,06	0,06	-0,03	-0,02	1,00										
nbloop	0,40	0,53	0,35	-0,08	0,06	0,06	0,02	0,01	-0,13	-0,12	0,06	1,00									
propuninv	-0,11	-0,10	-0,03	0,12	0,05	0,07	0,22	0,21	0,02	0,01	-0,15	0,02	1,00								
excl_nano_pat	0,19	0,04	0,22	0,06	0,03	0,04	0,01	0,01	0,00	0,00	-0,02	0,08	-0,03	1,00							
excl_bio_pat	0,04	0,09	-0,06	0,10	0,08	0,08	0,04	0,05	0,05	0,03	0,02	0,00	0,03	0,06	1,00						
ltotcontmoyan3wu	-0,04	-0,03	-0,06	-0,02	0,14	0,14	0,20	0,19	-0,14	-0,13	-0,06	0,02	0,08	0,02	0,00	1,00					
ltotsubvmoyan3wu	-0,05	-0,06	-0,16	0,03	0,06	0,06	0,29	0,29	-0,21	-0,18	-0,04	-0,05	-0,04	0,00	-0,03	0,57	1,00				
nbartmoyan3wu	-0,04	-0,03	-0,06	-0,02	0,14	0,14	0,20	0,19	-0,14	-0,13	-0,06	0,02	0,08	0,02	0,00	1,00	0,57	1,00			
codechair33	-0,07	-0,08	0,01	0,08	0,15	0,15	0,24	0,25	0,00	-0,01	0,02	-0,03	0,14	0,02	0,03	0,10	0,03	0,10	1,00		
age	-0,06	-0,05	-0,17	0,19	0,08	0,08	0,25	0,26	-0,09	-0,08	0,06	-0,14	0,00	-0,02	0,05	0,14	0,35	0,14	0,09	1,00	
sq_age	-0,07	-0,05	-0,18	0,18	0,07	0,07	0,24	0,25	-0,09	-0,08	0,05	-0,12	-0,02	-0,03	0,04	0,13	0,36	0,13	0,09	0,98	1,00

Tableau M-2: table de corrélation entre des variables pour les blocs D, E et F

	nbclaims0	nbpat0	nbctip50	nbartmoyan3	ltotcontmoyan3	sq_ltotcontmoyan3	ltotsubvmoyan3	sq_ltotsubvmoyan3	lmcliquind23	sq_lmcliquind23	deg_app0	nbloop0	propuninv0	nano_excl	bio_excl	ltotcontmoyan3wuu	ltotsubvmoyan3wuu	nbartmoyan3wuu	codechair33	age	sq_age
nbclaims0	1,00																				
nbpat0	0,82	1,00																			
nbctip50	0,60	0,62	1,00																		
nbartmoyan3	0,08	0,09	0,06	1,00																	
ltotcontmoyan3	0,06	0,04	0,04	0,10	1,00																
sq_ltotcontmoyan3	0,07	0,05	0,04	0,12	0,98	1,00															
ltotsubvmoyan3	0,03	0,03	0,01	0,25	0,28	0,30	1,00														
sq_ltotsubvmoyan3	0,03	0,03	0,01	0,26	0,29	0,31	0,99	1,00													
lmcliquind23	0,04	0,06	0,04	0,22	0,01	0,02	-0,09	-0,08	1,00												
sq_lmcliquind23	0,03	0,05	0,04	0,17	0,01	0,02	-0,09	-0,08	0,95	1,00											
deg_app0	0,03	0,04	0,03	0,36	-0,01	0,00	0,13	0,13	0,11	0,07	1,00										
nbloop0	0,52	0,62	0,46	0,03	0,05	0,06	0,03	0,02	0,00	-0,01	0,03	1,00									
propuninv0	0,47	0,52	0,36	0,14	0,09	0,10	0,08	0,08	0,05	0,04	0,06	0,35	1,00								
nano_excl	0,01	-0,01	0,00	-0,07	-0,03	-0,03	-0,03	-0,03	-0,07	-0,04	-0,07	0,02	-0,02	1,00							
bio_excl	-0,06	-0,06	-0,06	-0,22	-0,01	-0,01	-0,10	-0,10	-0,04	-0,04	-0,14	-0,04	-0,07	-0,14	1,00						
ltotcontmoyan3wuu	0,00	0,00	-0,01	-0,02	0,14	0,14	0,20	0,19	-0,14	-0,13	-0,02	0,01	0,02	-0,01	0,00	1,00					
ltotsubvmoyan3wuu	-0,02	-0,02	-0,04	0,03	0,06	0,06	0,29	0,29	-0,21	-0,18	0,00	-0,02	-0,02	-0,02	0,01	0,57	1,00				
nbartmoyan3wuu	0,00	0,00	-0,01	-0,02	0,14	0,14	0,20	0,19	-0,14	-0,13	-0,02	0,01	0,02	-0,01	0,00	1,00	0,57	1,00			
codechair33	0,00	0,01	0,02	0,08	0,15	0,15	0,24	0,25	0,00	-0,01	0,07	0,00	0,05	-0,03	-0,10	0,10	0,03	0,10	1,00		
age	0,02	0,02	-0,01	0,19	0,08	0,08	0,25	0,26	-0,09	-0,08	0,08	-0,01	0,04	-0,08	-0,12	0,14	0,35	0,14	0,09	1,00	
sq_age	0,01	0,01	-0,02	0,18	0,07	0,07	0,24	0,25	-0,09	-0,08	0,07	-0,01	0,03	-0,07	-0,11	0,13	0,36	0,13	0,09	0,98	1,00

## ANNEXE N – Régressions de deuxième étape du bloc A

Tableau N-1: Régressions à binomiale négative (deuxième étape) du bloc A – fait de breveter et nombre de brevets

Variable	nbreg du Bloc A - fait de breveter et nombre de brevets													
	nbreg_vce		nbreg_art		nbreg_vce1		nbreg_subv		nbreg_subv1		nbreg_subv2		nbreg_vce2	nbreg_cont
Endogénéité			x				x		x		x			x
L1.nbarticlemoyan3	0.0073 (0.0248)		0.0240 (0.0403)										0.0150 (0.0249)	0.0139 (0.0248)
L2.ltotcontmoyan3	0.0324 (0.0113)	***	0.0317 (0.0112)	***									0.0414 (0.0121)	***
L1.ltotcontmoyan3					0.0382 (0.0120)	***	0.0372 (0.0151)	**	0.0374 (0.0152)	**	0.0377 (0.0151)	**		0.0824 (0.0281)
L2.lmcliquind23	0.3866 (0.1326)	***	0.3713 (0.1328)	***	0.3880 (0.1246)	***	0.3890 (0.1184)	***	0.3887 (0.1183)	***	0.3883 (0.1178)	***	0.3687 (0.1303)	0.4664 (0.1392)
L2.sq_lmcliquind23	-0.0152 (0.0053)	***	-0.0146 (0.0053)	***	-0.0153 (0.0049)	***	-0.0153 (0.0047)	***	-0.0153 (0.0047)	***	-0.0153 (0.0047)	***	-0.0145 (0.0052)	-0.0183 (0.0055)
L1.deg_app0	0.0699 (0.0468)		0.0623 (0.0513)		0.0790 (0.0458)	*	0.0782 (0.0455)	*	0.0784 (0.0453)	*	0.0787 (0.0459)	*	0.0745 (0.0469)	0.0798 (0.0483)
L1.avg_pat_past	0.5299 (0.1897)	***	0.5221 (0.1943)	***	0.5266 (0.1844)	***	0.5267 (0.1883)	***	0.5269 (0.1881)	***	0.5270 (0.1888)	***	0.5274 (0.1813)	***
L1.o_patpast	1.0470 (0.0773)	***	1.0394 (0.0795)	***	1.0583 (0.0752)	***	1.0560 (0.0766)	***	1.0562 (0.0767)	***	1.0566 (0.0762)	***	1.0559 (0.0769)	1.2258 (0.0825)
d1997	-0.6102 (0.2167)	***	-0.6022 (0.2171)	***	-0.5424 (0.2200)	**	-0.5262 (0.3259)		-0.5298 (0.3270)		-0.5353 (0.3254)	*	-0.6002 (0.2180)	-0.4071 (0.2250)
d1998	-0.4422 (0.2656)	*	-0.4505 (0.2641)	*	-0.4395 (0.2664)	*	-0.4425 (0.2631)	*	-0.4427 (0.2632)	*	-0.4429 (0.2627)	*	-0.4273 (0.2643)	-0.2930 (0.2706)
d1999	-0.3283 (0.2227)		-0.3327 (0.2229)		-0.2818 (0.2274)		-0.2867 (0.2301)		-0.2862 (0.2302)		-0.2853 (0.2303)		-0.2525 (0.2411)	-0.1481 (0.2473)
d2000	-0.3093 (0.2018)		-0.3152 (0.2016)		-0.2447 (0.2048)		-0.2496 (0.2049)		-0.2492 (0.2050)		-0.2484 (0.2048)		-0.2171 (0.2090)	-0.0924 (0.2200)
d2001	-0.7158 (0.2437)	***	-0.7254 (0.2438)	***	-0.6543 (0.2443)	***	-0.6602 (0.2448)	***	-0.6596 (0.2450)	***	-0.6588 (0.2446)	***	-0.6463 (0.2502)	-0.4985 (0.2564)
d2002	-0.3153 (0.2725)		-0.3205 (0.2725)		-0.2428 (0.2709)		-0.2482 (0.2865)		-0.2476 (0.2872)		-0.2465 (0.2856)		-0.2275 (0.2746)	-0.0572 (0.2786)
d2003	-0.7961 (0.2525)	***	-0.8054 (0.2525)	***	-0.6964 (0.2533)	***	-0.7024 (0.2656)	***	-0.7018 (0.2660)	***	-0.7009 (0.2657)	***	-0.6918 (0.2609)	-0.4976 (0.2659)
d2004	-1.2443	***	-1.2534	***	-1.1518	***	-1.1592	***	-1.1584	***	-1.1571	***	-1.1332	-0.9592

nbreg du Bloc A - fait de breveter et nombre de brevets																
Variable	nbreg_vce		nbreg_art		nbreg_vce1		nbreg_subv		nbreg_subv1		nbreg_subv2		nbreg_vce2		nbreg_cont	
Endogénéité			x				x				x				x	
	(0.2706)		(0.2698)		(0.2750)		(0.2849)		(0.2853)		(0.2848)		(0.2799)		(0.2864)	
d2005	-1.6824	***	-1.6939	***	-1.5870	***	-1.5951	***	-1.5941	***	-1.5926	***	-1.5799	***	-1.4202	***
	(0.3631)		(0.3577)		(0.3704)		(0.3719)		(0.3728)		(0.3712)		(0.3682)		(0.3674)	
d01Laval	-0.2468		-0.2364		-0.2669		-0.2715		-0.2707		-0.2694		-0.2010		-0.2799	
	(0.2468)		(0.2444)		(0.2401)		(0.2071)		(0.2065)		(0.2076)		(0.2398)		(0.2521)	
d04MtlG	-0.1734		-0.1638		-0.1908		-0.1950		-0.1942		-0.1930		-0.1482		-0.1313	
	(0.2485)		(0.2473)		(0.2400)		(0.2042)		(0.2040)		(0.2036)		(0.2433)		(0.2646)	
d07Concordia	-0.5972		-0.5808		-0.6381		-0.6254		-0.6258		-0.6263		-0.6329		-0.5876	
	(0.7441)		(0.7422)		(0.7389)		(0.7460)		(0.7461)		(0.7460)		(0.7467)		(0.7499)	
d09UQAMG	-0.4432		-0.4304		-0.4922	*	-0.4933	*	-0.4933	*	-0.4931	*	-0.4440		-0.4920	*
	(0.2885)		(0.2866)		(0.2826)		(0.2805)		(0.2801)		(0.2812)		(0.2868)		(0.2965)	
d08SherbrookeG	-0.1955		-0.1836		-0.2254		-0.2258		-0.2250		-0.2239		-0.1524		-0.2158	
	(0.3166)		(0.3148)		(0.3117)		(0.2913)		(0.2909)		(0.2912)		(0.3138)		(0.3267)	
L_nbart_res			-0.0278													
			(0.0471)													
L1.ltotsubvmoyan3					0.0581		0.0621		0.0605		0.0580					
					(0.0589)		(0.1174)		(0.1179)		(0.1158)					
L2.ltotsubvmoyan3													0.0316		0.0839	
													(0.0570)		(0.0639)	
L1.sq_ltotsubvmoyan3					-0.0048		-0.0048		-0.0047		-0.0047					
					(0.0044)		(0.0045)		(0.0045)		(0.0045)					
L2.sq_ltotsubvmoyan3													-0.0047		-0.0098	*
													(0.0046)		(0.0053)	
L_subv_res							-0.0070									
							(0.1003)									
L_subv_res1									-0.0055							
									(0.1009)							
L_subv_res2											-0.0033					
											(0.1000)					
L_cont_res															-0.0412	
															(0.0262)	
_cons	-3.6259	***	-3.6058	***	-3.7469	***	-3.7839	***	-3.7711	***	-3.7520	***	-3.5361	***	-4.0387	***
	(0.3830)		(0.3834)		(0.4162)		(1.0086)		(1.0128)		(1.0074)		(0.3715)		(0.4585)	
Inalpha	1.4052	***	1.4070	***	1.3902	***	1.3893	***	1.3893	***	1.3893	***	1.3900	***	1.5218	***
	(0.1992)		(0.2002)		(0.2016)		(0.2006)		(0.2008)		(0.1998)		(0.2001)		(0.1805)	
Statistics																
N	8793		8772		8786		8769		8769		8769		8789		8771	

nbreg du Bloc A - fait de breveter et nombre de brevets								
Variable	nbreg_vce	nbreg_art	nbreg_vce1	nbreg_subv	nbreg_subv1	nbreg_subv2	nbreg_vce2	nbreg_cont
Endogénéité		x		x	x	x		x
df_m	21	22	22	23	23	23	23	23
k	23	24	24	25	25	25	25	25
k_aux	1	1	1	1	1	1	1	1
k_eq	2	2	2	2	2	2	2	2
k_eq_model	1	1	1	1	1	1	1	1
k_dv	1	1	1	1	1	1	1	1
ll	-2202.98	-2201.8	-2200.87	-2200.11	-2200.11	-2200.11	-2200.38	-2217.29
ll_0	-2531.47	-2530.18	-2531.04	-2530	-2530	-2530	-2531.23	-2530.12
chi2	537.5915	541.0756	554.194	553.7615	553.6918	553.5195	541.6247	493.1793
aic	4451.968	4451.609	4449.749	4450.213	4450.218	4450.223	4450.765	4484.574
bic	4614.847	4621.512	4619.691	4627.188	4627.192	4627.197	4627.797	4661.554
p	2E-100	1.9E-100	3.5E-103	2.2E-102	2.3E-102	2.4E-102	7.5E-100	9.26E-90

Tableau N-2: Régressions à binomiale négative en panel (deuxième étape) du bloc A - fait de breveter et nombre de brevets

xtnbreg du Bloc A - fait de breveter et nombre de brevets																
Variable	xtnbreg_0		xtnbreg_art		xtnbreg_1		xtnbreg_subv		xtnbreg_subv1		xtnbreg_subv2		xtnbreg_2		xtnbreg_cont	
Endogénéité	x		x		x		x		x		x		x		x	
L1.nbarticlemoyan3	0.0321	*	0.0844										0.0368	*	0.0333	*
	(0.0194)		(0.0544)										(0.0199)		(0.0200)	
L2.ltotcontmoyan3	0.0259	***	0.0243	**									0.0320	***		
	(0.0098)		(0.0099)										(0.0105)			
L1.ltotcontmoyan3					0.0435	***	0.0302	**	0.0302	**	0.0307	**			0.0888	**
					(0.0103)		(0.0121)		(0.0121)		(0.0121)				(0.0434)	
L2.lmcliquind23	0.2414	**	0.2088	*	0.2775	***	0.2792	***	0.2793	***	0.2800	***	0.2346	**	0.2402	**
	(0.1060)		(0.1111)		(0.1026)		(0.1032)		(0.1032)		(0.1031)		(0.1062)		(0.1059)	
L2.sq_lmcliquind23	-0.0093	**	-0.0079	*	-0.0108	***	-0.0108	***	-0.0108	***	-0.0108	***	-0.0090	**	-0.0092	**
	(0.0042)		(0.0044)		(0.0040)		(0.0040)		(0.0040)		(0.0040)		(0.0042)		(0.0042)	
L1.deg_app0	0.0876	***	0.0714	**	0.1065	***	0.1000	***	0.1001	***	0.1003	***	0.0911	***	0.0917	***
	(0.0327)		(0.0364)		(0.0319)		(0.0321)		(0.0321)		(0.0321)		(0.0328)		(0.0330)	
L1.avg_pat_past	0.2023	***	0.2071	***	0.2161	***	0.1909	***	0.1908	***	0.1919	***	0.2001	***	0.1919	***
	(0.0531)		(0.0536)		(0.0544)		(0.0561)		(0.0561)		(0.0561)		(0.0528)		(0.0543)	
L1.o_patpast	0.8137	***	0.7849	***	0.8444	***	0.8329	***	0.8329	***	0.8334	***	0.8219	***	0.8136	***
	(0.0744)		(0.0793)		(0.0726)		(0.0731)		(0.0731)		(0.0731)		(0.0741)		(0.0748)	
d1997	-0.4422	**	-0.4474	**	-0.4071	*	-0.1451		-0.1452		-0.1545		-0.4012	*	-0.3442	
	(0.2137)		(0.2135)		(0.2164)		(0.2502)		(0.2502)		(0.2505)		(0.2152)		(0.2162)	
d1998	-0.3875	*	-0.4090	*	-0.3611	*	-0.3937	*	-0.3940	*	-0.3917	*	-0.3542	*	-0.3727	*
	(0.2099)		(0.2106)		(0.2087)		(0.2091)		(0.2091)		(0.2091)		(0.2109)		(0.2125)	
d1999	-0.2096		-0.2322		-0.1894		-0.2652		-0.2652		-0.2634		-0.1213		-0.1103	
	(0.1990)		(0.1998)		(0.1972)		(0.2001)		(0.2001)		(0.2002)		(0.2049)		(0.2059)	
d2000	-0.1558		-0.1729		-0.1213		-0.1881		-0.1882		-0.1867		-0.0624		-0.0655	
	(0.2043)		(0.2045)		(0.2026)		(0.2048)		(0.2049)		(0.2049)		(0.2104)		(0.2127)	
d2001	-0.6378	***	-0.6629	***	-0.5650	**	-0.6487	***	-0.6489	***	-0.6457	***	-0.5388	**	-0.5248	**
	(0.2253)		(0.2264)		(0.2231)		(0.2264)		(0.2264)		(0.2264)		(0.2309)		(0.2333)	
d2002	-0.4576	**	-0.4860	**	-0.3797	*	-0.4896	**	-0.4898	**	-0.4856	**	-0.3531		-0.3395	
	(0.2250)		(0.2266)		(0.2232)		(0.2289)		(0.2289)		(0.2289)		(0.2313)		(0.2334)	
d2003	-0.7843	***	-0.8126	***	-0.6907	***	-0.8140	***	-0.8141	***	-0.8097	***	-0.6702	***	-0.6330	***
	(0.2318)		(0.2332)		(0.2305)		(0.2376)		(0.2376)		(0.2376)		(0.2391)		(0.2404)	
d2004	-1.2415	***	-1.2676	***	-1.1553	***	-1.2994	***	-1.2996	***	-1.2943	***	-1.1279	***	-1.0941	***
	(0.2438)		(0.2448)		(0.2434)		(0.2522)		(0.2522)		(0.2522)		(0.2510)		(0.2519)	
d2005	-1.7875	***	-1.8109	***	-1.6845	***	-1.8552	***	-1.8555	***	-1.8493	***	-1.6743	***	-1.6218	***
	(0.2731)		(0.2737)		(0.2730)		(0.2842)		(0.2842)		(0.2843)		(0.2797)		(0.2803)	
d01Laval	0.0877		0.1156		0.0406		-0.0027		-0.0028		-0.0012		0.1296		0.0151	



xtnbreg du Bloc A - fait de breveter et nombre de brevets																
Variable	xtnbreg_0		xtnbreg_art		xtnbreg_1		xtnbreg_subv		xtnbreg_subv1		xtnbreg_subv2		xtnbreg_2		xtnbreg_cont	
Endogénéité			x				x		x		x				x	
	(0.1694)		(0.1724)		(0.1688)		(0.1705)		(0.1705)		(0.1704)		(0.1709)		(0.1918)	
d04MtIG	0.1066		0.1329		0.0486		0.0053		0.0053		0.0062		0.1318		0.0349	
	(0.1606)		(0.1632)		(0.1598)		(0.1617)		(0.1617)		(0.1617)		(0.1613)		(0.1740)	
d07Concordia	-0.7878		-0.7753		-0.8204	*	-0.7463		-0.7463		-0.7487		-0.8141		-0.7474	
	(0.4949)		(0.4955)		(0.4951)		(0.4974)		(0.4973)		(0.4973)		(0.4964)		(0.4973)	
d09UQAMG	-0.0828		-0.0443		-0.1687		-0.1661		-0.1662		-0.1661		-0.0709		-0.1978	
	(0.1990)		(0.2033)		(0.1987)		(0.2001)		(0.2001)		(0.2000)		(0.2003)		(0.2197)	
d08SherbrookeG	0.0685		0.0999		0.0265		-0.0244		-0.0244		-0.0230		0.1137		0.0538	
	(0.2349)		(0.2373)		(0.2344)		(0.2370)		(0.2370)		(0.2370)		(0.2360)		(0.2423)	
L_nbart_xt_ue			-0.0668													
			(0.0651)													
L1.ltotsubvmoyan3					0.0414		0.1616	**	0.1618	**	0.1558	**				
					(0.0536)		(0.0796)		(0.0796)		(0.0793)					
L2.ltotsubvmoyan3													-0.0076		0.0129	
													(0.0496)		(0.0511)	
L1.sq_ltotsubvmoyan3					-0.0044		-0.0058		-0.0058		-0.0057					
					(0.0040)		(0.0041)		(0.0041)		(0.0041)					
L2.sq_ltotsubvmoyan3													-0.0014		-0.0041	
													(0.0041)		(0.0044)	
L_subv_xt_ue							-0.1146	**								
							(0.0544)									
L_subv_xt_ue1									-0.1146	**						
									(0.0544)							
L_subv_xt_ue2											-0.1101	**				
											(0.0545)					
L_cont_xt_ue															-0.0458	
															(0.0438)	
_cons	-2.1139	***	-2.0995	***	-2.1820	***	-3.0658	***	-3.0664	***	-3.0322	***	-2.0167	***	-2.1062	***
	(0.4033)		(0.4036)		(0.4382)		(0.6171)		(0.6170)		(0.6175)		(0.4073)		(0.4066)	
ln_r																
_cons	1.6728	***	1.6711	***	1.6853	***	1.6721	***	1.6721	***	1.6724	***	1.6837	***	1.6872	***
	(0.1465)		(0.1464)		(0.1470)		(0.1470)		(0.1470)		(0.1470)		(0.1464)		(0.1464)	
ln_s																
_cons	0.0554		0.0546		0.0635		0.0504		0.0502		0.0512		0.0529		0.0623	
	(0.2092)		(0.2084)		(0.2074)		(0.2072)		(0.2072)		(0.2071)		(0.2082)		(0.2059)	
Statistics																
N	8793		8772		8786		8769		8769		8769		8789		8771	

xtnbreg du Bloc A - fait de breveter et nombre de brevets								
Variable	xtnbreg_0	xtnbreg_art	xtnbreg_1	xtnbreg_subv	xtnbreg_subv1	xtnbreg_subv2	xtnbreg_2	xtnbreg_cont
Endogénéité	x		x		x		x	
N_g	1460	1460	1460	1460	1460	1460	1460	1460
df_m	21	22	22	23	23	23	23	24
k	24	25	25	26	26	26	26	27
k_aux	2	2	2	2	2	2	2	2
k_eq	3	3	3	3	3	3	3	3
k_eq_model	1	1	1	1	1	1	1	1
k_dv	1	1	1	1	1	1	1	1
ll	-2125.19	-2123.87	-2120.76	-2117.89	-2117.89	-2118.07	-2123.14	-2117.15
ll_c	-2198.17	-2197.22	-2191.71	-2188.95	-2188.95	-2189.06	-2194.47	-2189.59
chi2	301.9432	304.1757	312.4414	315.4662	315.4361	315.3057	307.492	320.3596
chi2_c	145.9525	146.7089	141.8941	142.1114	142.1303	141.9822	142.6636	144.884
aic	4298.389	4297.739	4291.516	4287.781	4287.773	4288.134	4298.273	4288.306
bic	4468.35	4474.722	4468.539	4471.834	4471.826	4472.188	4482.386	4479.444
g_min	1	1	1	1	1	1	1	1
g_avg	6.022603	6.008219	6.017808	6.006164	6.006164	6.006164	6.019863	6.007534
g_max	10	10	10	10	10	10	10	10
r	5.326955	5.318107	5.394026	5.323105	5.323104	5.325174	5.3857	5.404474
s	1.056932	1.056121	1.065557	1.051664	1.051493	1.052544	1.054331	1.064323
p	1.28E-51	1.74E-51	3.63E-53	3.39E-53	3.43E-53	3.65E-53	1.4E-51	1.3E-53

Tableau N-3: Régressions à binomiale négative à zéro-augmenté (deuxième étape) du bloc A – fait de breveter et nombre de brevets

zinb du Bloc A – fait de breveter et nombre de brevets						
Variable	zinb_vce	zinb_art	zinb_subv	zinb_subv1	zinb_subv2	zinb_cont
Endogénéité		x	x	x	x	x
L_nbarticlemoyan3 <sup>57</sup>	0.0268 (0.0294)					0.0231 (0.0261)
L2_ltotsubvmoyan3	0.0093 (0.0941)					0.0320 (0.0957)
L2_sq_ltotsubvmoyan3	-0.0036 (0.0070)					-0.0064 (0.0071)
L2_ltotcontmoyan3	0.0001 (0.0260)	-0.0059 (0.0160)				
L2_lmliquind23	0.1649 (0.1964)	0.2449 (0.2136)	0.3600 * (0.2067)	0.3600 * (0.2067)	0.3614 * (0.2065)	0.2421 (0.1927)
L2_sq_lmliquind23	-0.0065 (0.0078)	-0.0098 (0.0084)	-0.0145 * (0.0082)	-0.0145 * (0.0082)	-0.0145 * (0.0082)	-0.0097 (0.0076)
L_deg_app0	-0.0447 (0.0797)	-0.1390 (0.0915)	-0.1251 (0.0855)	-0.1249 (0.0853)	-0.1241 (0.0859)	-0.1409 * (0.0796)
L_avg_pat_past	0.4439 *** (0.0667)	0.3584 *** (0.0732)	0.3547 *** (0.0661)	0.3551 *** (0.0659)	0.3544 *** (0.0674)	0.3945 *** (0.0492)
L_o_patpast	0.3664 *** (0.1163)	0.4695 *** (0.1158)	0.4748 *** (0.1088)	0.4747 *** (0.1089)	0.4771 *** (0.1086)	0.5322 *** (0.1141)
L_nbloop0	0.3647 *** (0.0507)	0.3502 *** (0.0559)	0.3394 *** (0.0558)	0.3393 *** (0.0559)	0.3385 *** (0.0556)	0.3257 *** (0.0492)
L_deg_app_subv	-0.0169 * (0.0094)					
L_deg_app_nbartcile	0.0036 (0.0083)					
d1997	-0.8304 *** (0.2230)	-0.8729 *** (0.2131)	-0.7484 ** (0.2931)	-0.7506 ** (0.2929)	-0.7622 ** (0.2980)	-0.8823 *** (0.2060)
d1998	-0.5311 * (0.2731)	-0.7094 *** (0.2539)	-0.6728 *** (0.2493)	-0.6732 *** (0.2496)	-0.6729 *** (0.2483)	-0.7112 *** (0.2530)
d1999	-0.3388 (0.2349)	-0.5999 *** (0.2155)	-0.5613 *** (0.2157)	-0.5614 *** (0.2159)	-0.5576 *** (0.2156)	-0.4823 ** (0.2175)
d2000	-0.1636 (0.1979)	-0.4640 ** (0.1874)	-0.4329 ** (0.1845)	-0.4331 ** (0.1848)	-0.4296 ** (0.1840)	-0.3338 * (0.1835)
d2001	-0.8044 *** (0.2576)	-1.0020 *** (0.2358)	-0.9461 *** (0.2378)	-0.9461 *** (0.2381)	-0.9425 *** (0.2373)	-0.9116 *** (0.2408)
d2002	-0.3520 (0.2576)	-0.5548 ** (0.2433)	-0.5040 ** (0.2550)	-0.5043 ** (0.2558)	-0.4992 ** (0.2534)	-0.4489 * (0.2444)
d2003	-0.8061 *** (0.2730)	-1.0462 *** (0.2411)	-0.9847 *** (0.2561)	-0.9846 *** (0.2566)	-0.9805 *** (0.2554)	-0.9278 *** (0.2520)
d2004	-1.1907 *** (0.2797)	-1.4456 *** (0.2581)	-1.3966 *** (0.2711)	-1.3965 *** (0.2717)	-1.3917 *** (0.2700)	-1.3409 *** (0.2636)
d2005	-1.5413 *** (0.3528)	-1.8154 *** (0.3336)	-1.7628 *** (0.3489)	-1.7625 *** (0.3500)	-1.7573 *** (0.3474)	-1.7449 *** (0.3442)
d01Laval	-0.1634 (0.1929)	-0.2437 (0.2072)	-0.3025 * (0.1821)	-0.3025 * (0.1817)	-0.2997 * (0.1813)	-0.0585 (0.1975)
d04MtlG	-0.0309 (0.1997)	-0.0701 (0.2076)	-0.1292 (0.1788)	-0.1289 (0.1789)	-0.1270 (0.1765)	0.0476 (0.1941)
d07Concordia	-0.4318 (0.7525)	-0.4888 (0.7219)	-0.4499 (0.7343)	-0.4502 (0.7343)	-0.4532 (0.7343)	-0.5310 (0.7226)
d09UQAMG	-0.2544 (0.2420)	-0.2299 (0.2411)	-0.2709 (0.2391)	-0.2712 (0.2388)	-0.2716 (0.2388)	-0.0825 (0.2467)
d08SherbrookeG	-0.0746	-0.1392	-0.1904	-0.1902	-0.1889	-0.0283

<sup>57</sup> Lors de l'utilisation des commandes de tobit et de zinb, nous devons créer nos propres variables de lag car ces régressions ne tolèrent pas de spécifications temporelles. À toutes fins pratiques cependant, L.nbcitp5=L.nbcitp5.

zinb du Bloc A – fait de breveter et nombre de brevets						
Variable	zinb_vce	zinb_art	zinb_subv	zinb_subv1	zinb_subv2	zinb_cont
Endogénéité		x	x	x	x	x
L_nbart_pre	(0.2797)	(0.2762)	(0.2615)	(0.2613)	(0.2606)	(0.2755)
		0.0409				
		(0.0388)				
L_ltotcontmoyan3			-0.0041	-0.0040	-0.0033	
			(0.0182)	(0.0183)	(0.0181)	
L_subv_pre			0.0321			
			(0.0819)			
L_subv1_pre				0.0315		
				(0.0821)		
L_subv2_pre					0.0835	
					(0.7488)	
L_cont_pre						-0.0163
						(0.0160)
_cons	-0.6789	-1.3300 *	-1.8450	-1.8386	-1.8007	-1.0351 *
	(0.6499)	(0.6850)	(1.1423)	(1.1431)	(1.1610)	(0.6197)
inflata						
L_nbarticlemoyan3	-0.0506					-0.0774
	(0.0866)					(0.0952)
L2_ltotsubvmoyan3	0.0916					0.2721 *
	(0.1713)					(0.1611)
L2_sq_ltotsubvmoyan3	-0.0122					-0.0291 **
	(0.0145)					(0.0130)
L2_ltotcontmoyan3	-0.0221	-0.0835 **				
	(0.0490)	(0.0328)				
L2_lmliquind23	-0.3342	-0.2241	-0.0869	-0.0866	-0.0832	-0.1923
	(0.3640)	(0.3615)	(0.3116)	(0.3117)	(0.3113)	(0.3416)
L2_sq_lmliquind23	0.0125	0.0081	0.0027	0.0027	0.0026	0.0069
	(0.0145)	(0.0141)	(0.0123)	(0.0123)	(0.0122)	(0.0134)
L_deg_app0	-0.2731 **	-0.3814 ***	-0.3700 ***	-0.3700 ***	-0.3697 ***	-0.3498 ***
	(0.1280)	(0.1176)	(0.1181)	(0.1181)	(0.1182)	(0.1232)
L_o_patpast	-1.6980 ***					
	(0.4321)					
L_deg_app_subv	-0.0045					
	(0.0131)					
L_deg_app_nbpat	-0.0191					
	(0.0146)					
L_avg_pat_past		-12.4942 *	-11.6157 **	-11.6256 **	-11.5554 **	-10.0436 ***
		6.952577	5.035695	5.048998	4.982757	3.207141
L_ltotcontmoyan3			-0.07596 ***	-0.07597 ***	-0.07589 ***	
			0.029053	0.029052	0.029062	
_cons	3.5401 ***	2.6337 ***	2.3602 ***	2.3596 ***	2.3501 ***	2.8526 ***
	(0.9486)	(0.9610)	(0.8536)	(0.8539)	(0.8557)	(0.9439)
Inalpha						
_cons	0.560125 **	0.616311 **	0.569491 **	0.569678 **	0.568307 **	0.50655 **
	0.234858	0.270709	0.250917	0.250739	0.251866	0.232106
Statistics						
N	8340	8772	8775	8775	8775	8774
df_m	26	22	22	22	22	24
k	39	30	30	30	30	34
k_aux	1	1	1	1	1	1
k_eq	3	3	3	3	3	3
k_eq_model	1	1	1	1	1	1
k_dv	1	1	1	1	1	1
ll	-2073.46	-2148.64	-2147.9	-2147.9	-2147.95	-2145.45
ll_0	-2220.88	-2298.43	-2294.93	-2294.93	-2294.93	-2302.91
chi2	314.4431	205.8846	210.6609	210.5657	210.8416	363.5811
aic	4224.913	4357.282	4355.796	4355.805	4355.901	4358.903
bic	4499.037	4569.662	4568.186	4568.195	4568.29	4599.607
p	2.7E-51	7.99E-32	9.21E-33	9.62E-33	8.49E-33	2.14E-62

Tableau N-4: Régressions tobit (deuxième étape) du bloc A – fait de breveter et nombre de brevets

tobit du Bloc A – fait de breveter et nombre de brevets									
Variable	HSM_tobit_art		HSM_tobit_subv		HSM_tobit_subv_1		HSM_tobit_subv_2		HSM_tobit_cont
Endogénéité	x		x		x		x		x
L_ttotcontmoyan3	0.0271	***	0.0180	*	0.0180	*	0.0187	*	-0.0268 *
	(0.0073)		(0.0096)		(0.0096)		(0.0095)		(0.0162)
L2_lmcliqind23	0.1776	**	0.2361	***	0.2361	***	0.2359	***	0.1723 **
	(0.0768)		(0.0733)		(0.0733)		(0.0731)		(0.0725)
L2_sq_lmcliqind23	-0.0068	**	-0.0091	***	-0.0091	***	-0.0091	***	-0.0066 **
	(0.0030)		(0.0029)		(0.0029)		(0.0029)		(0.0029)
L_deg_app0	0.0531	**	0.0612	***	0.0613	***	0.0620	***	0.0528 **
	(0.0250)		(0.0229)		(0.0229)		(0.0229)		(0.0225)
L_avg_pat_past	0.3545	***	0.3641	***	0.3644	***	0.3642	***	0.4384 ***
	(0.0873)		(0.0705)		(0.0708)		(0.0691)		(0.0546)
L_o_patpast	0.6947	***	0.7037	***	0.7036	***	0.7050	***	0.7057 ***
	(0.0510)		(0.0485)		(0.0485)		(0.0485)		(0.0475)
L_nbloop0	0.3078	***	0.3038	***	0.3038	***	0.3026	***	0.3139 ***
	(0.0695)		(0.0697)		(0.0698)		(0.0696)		(0.0676)
L_nbarticlemoyan3	0.0381								0.0380 ***
	(0.0279)								(0.0145)
L_nbart_res	-0.0155								
	(0.0339)								
d1997	-0.3230	**	-0.0592		-0.0604		-0.0760		-0.3373 **
	(0.1620)		(0.2006)		(0.2006)		(0.2013)		(0.1630)
d1998	-0.3174	**	-0.2935	*	-0.2936	*	-0.2942	*	-0.2913 *
	(0.1598)		(0.1602)		(0.1602)		(0.1602)		(0.1609)
d1999	-0.1753		-0.1925		-0.1923		-0.1911		-0.1288
	(0.1596)		(0.1602)		(0.1602)		(0.1603)		(0.1651)
d2000	-0.1074		-0.1106		-0.1105		-0.1098		-0.0440
	(0.1578)		(0.1588)		(0.1588)		(0.1588)		(0.1610)
d2001	-0.4113	**	-0.4090	**	-0.4088	**	-0.4079	**	-0.3566 **
	(0.1782)		(0.1787)		(0.1788)		(0.1786)		(0.1818)
d2002	-0.2891	*	-0.2966	*	-0.2967	*	-0.2941	*	-0.2392
	(0.1735)		(0.1753)		(0.1753)		(0.1753)		(0.1767)
d2003	-0.4605	***	-0.4635	**	-0.4634	**	-0.4622	**	-0.4166 **
	(0.1776)		(0.1812)		(0.1812)		(0.1812)		(0.1826)
d2004	-0.7593	***	-0.7760	***	-0.7758	***	-0.7744	***	-0.7257 ***
	(0.1838)		(0.1889)		(0.1890)		(0.1889)		(0.1868)
d2005	-1.1903	***	-1.2112	***	-1.2109	***	-1.2089	***	-1.1725 ***
	(0.1865)		(0.1917)		(0.1918)		(0.1916)		(0.1901)
d01Laval	0.0631		-0.0158		-0.0161		-0.0125		0.2200 *
	(0.1151)		(0.1132)		(0.1131)		(0.1134)		(0.1174)
d04MtlG	0.0650		-0.0146		-0.0144		-0.0127		0.1668
	(0.1153)		(0.1101)		(0.1101)		(0.1101)		(0.1118)
d07Concordia	-0.3896		-0.3929		-0.3930		-0.3952		-0.4316
	(0.4889)		(0.4861)		(0.4861)		(0.4858)		(0.4876)
d09UQAMG	-0.1003		-0.1504		-0.1510		-0.1493		0.0402
	(0.1425)		(0.1434)		(0.1434)		(0.1434)		(0.1437)
d08SherbrookeG	0.0659		-0.0170		-0.0169		-0.0146		0.1633
	(0.1766)		(0.1781)		(0.1782)		(0.1780)		(0.1762)
L_ttotsubvmoyan3			0.1379	**	0.1376	**	0.1289	**	
			(0.0635)		(0.0636)		(0.0626)		
L_sq_ttotsubvmoyan3			-0.0039		-0.0039		-0.0036		
			(0.0028)		(0.0028)		(0.0028)		
L_subv_res			-0.1017	**					
			(0.0480)						
L_subv_res1					-0.1013	**			
					(0.0479)				
L_subv_res2							-0.0954	**	
							(0.0483)		
L2_ttotsubvmoyan3									-0.0246
									(0.0383)
L2_sq_ttotsubvmoyan3									0.0022
									(0.0033)
L_cont_res									0.0614 ***
									(0.0164)
_cons	-3.0964	***	-4.0720	***	-4.0683	***	-4.0164	***	-3.0111 ***

tobit du Bloc A – fait de breveter et nombre de brevets										
Variable	HSM_tobit_art		HSM_tobit_subv		HSM_tobit_subv_1		HSM_tobit_subv_2		HSM_tobit_cont	
Endogénéité	x		x		x		x		x	
	(0.2754)		(0.5079)		(0.5075)		(0.5081)		(0.2730)	
sigma										
_cons	1.4818	***	1.4803	***	1.4803	***	1.4806	***	1.4740	***
	(0.0369)		(0.0378)		(0.0378)		(0.0378)		(0.0376)	
Statistics										
N	8769		8769		8769		8769		8771	
df_m	23		24		24		24		25	
k_aux	1		1		1		1		1	
ll	-2066.95		-2065.75		-2065.77		-2066.11		-2063.01	
ll_0	-2405.17		-2405.17		-2405.17		-2405.17		-2405.3	
aic	4183.896		4183.496		4183.533		4184.224		4180.013	
bic	4360.871		4367.55		4367.586		4368.277		4371.152	
r2_p	0.140623		0.141121		0.141114		0.14097		0.142306	
p	3.3E-129		2.4E-135		4.3E-135		5.9E-136		1.7E-154	

## ANNEXE O – Régressions de deuxième étape du bloc B

Tableau O-1: Régressions à binomiale négative (deuxième étape) du bloc B – breveter avec revendications

Nbreg du Bloc B – breveter avec revendications																
Variable	nbreg_vce		nbreg_art		nbreg_vce1		nbreg_subv		nbreg_subv1		nbreg_subv2		nbreg_vce2		nbreg_cont	
Endogénéité	x		x		x		x		x		x		x		x	
L1.nbarticlemoyan3	0.0456		0.1200	**												
	(0.0408)		(0.0542)													
L1.ltotcontmoyan3	0.0518	***	0.0410	**	0.0614	***	0.0405	*	0.0407	*	0.0395	*	0.0614	***	0.0857	
	(0.0159)		(0.0161)		(0.0174)		(0.0230)		(0.0221)		(0.0235)		(0.0174)		(0.1589)	
L1.lmliquind23	0.1637		0.0870		0.1767		0.2207		0.2308		0.2229		0.1767		0.1689	
	(0.1757)		(0.1783)		(0.1666)		(0.1695)		(0.1680)		(0.1701)		(0.1666)		(0.1693)	
L1.sq_lmliquind23	-0.0068		-0.0035		-0.0074		-0.0090		-0.0094		-0.0091		-0.0074		-0.0071	
	(0.0069)		(0.0070)		(0.0065)		(0.0066)		(0.0065)		(0.0066)		(0.0065)		(0.0066)	
deg_app0	0.1246	**	0.0996	*	0.1581	***	0.1448	**	0.1418	**	0.1446	**	0.1581	***	0.1623	***
	(0.0570)		(0.0566)		(0.0546)		(0.0563)		(0.0562)		(0.0564)		(0.0546)		(0.0574)	
L1.avg_pat_past	5.3042	***	5.2067	***	5.5053	***	5.3894	***	5.3915	***	5.3889	***	5.5053	***	5.4751	***
	(0.8278)		(0.8071)		(0.8395)		(0.8517)		(0.8494)		(0.8487)		(0.8395)		(0.8327)	
nano_excl	-1.1939	***	-1.3635	***	-1.1853	***	-1.1923	**	-0.9761	**	-1.1930	**	-1.1853	***	-1.3421	**
	(0.4330)		(0.4866)		(0.4393)		(0.5213)		(0.4665)		(0.5201)		(0.4393)		(0.5316)	
bio_excl	-0.4385	**	-0.3619	*	-0.5649	***	-0.5085	**	-0.5362	**	-0.5043	**	-0.5649	***	-0.5388	**
	(0.2109)		(0.2149)		(0.2068)		(0.2089)		(0.2086)		(0.2094)		(0.2068)		(0.2101)	
d1997	-0.5353		-0.5302		-0.4892		-0.1088		-0.0017		-0.1014		-0.4892		-0.5217	
	(0.3929)		(0.3888)		(0.4173)		(0.5682)		(0.5584)		(0.5681)		(0.4173)		(0.4176)	
d1998	-0.4305		-0.4965		-0.4547		-0.4166		-0.3656		-0.4158		-0.4547		-0.5009	
	(0.3541)		(0.3569)		(0.3419)		(0.3516)		(0.3503)		(0.3498)		(0.3419)		(0.3590)	
d1999	-0.5883	*	-0.7197	**	-0.4903		-0.6092	*	-0.5088		-0.6034	*	-0.4903		-0.6128	*
	(0.3382)		(0.3346)		(0.3334)		(0.3367)		(0.3387)		(0.3357)		(0.3334)		(0.3459)	
d2000	-0.5116		-0.5913	*	-0.4338		-0.4763		-0.4554		-0.4739		-0.4338		-0.4730	
	(0.3339)		(0.3326)		(0.3303)		(0.3286)		(0.3282)		(0.3278)		(0.3303)		(0.3483)	
d2001	-0.8735	**	-0.9746	***	-0.8561	**	-0.9374	**	-0.8973	**	-0.9416	**	-0.8561	**	-0.9285	**
	(0.3701)		(0.3737)		(0.3660)		(0.3660)		(0.3673)		(0.3661)		(0.3660)		(0.3796)	
d2002	-0.5187		-0.5376		-0.4206		-0.5059		-0.4564		-0.4997		-0.4206		-0.4909	
	(0.3764)		(0.3742)		(0.3716)		(0.3713)		(0.3722)		(0.3708)		(0.3716)		(0.3761)	
d2003	-0.9336	***	-0.9916	***	-0.8924	**	-0.9885	***	-0.9368	***	-0.9917	***	-0.8924	**	-0.9559	***
	(0.3555)		(0.3524)		(0.3496)		(0.3507)		(0.3515)		(0.3502)		(0.3496)		(0.3488)	
d2004	-1.3460	***	-1.4681	***	-1.2833	***	-1.3945	***	-1.3383	***	-1.3973	***	-1.2833	***	-1.3481	***
	(0.3923)		(0.3835)		(0.3904)		(0.3835)		(0.3865)		(0.3835)		(0.3904)		(0.3871)	
d2005	-1.7211	***	-1.8629	***	-1.6919	***	-1.8600	***	-1.8242	***	-1.8651	***	-1.6919	***	-1.7511	***

Nbreg du Bloc B – breveter avec revendications																
Variable	nbreg_vce		nbreg_art		nbreg_vce1		nbreg_subv		nbreg_subv1		nbreg_subv2		nbreg_vce2		nbreg_cont	
Endogénéité	x		x		x		x		x		x		x		x	
	(0.4850)		(0.4703)		(0.4618)		(0.4587)		(0.4571)		(0.4581)		(0.4618)		(0.4616)	
d01Laval	0.2049		0.3078		0.1465		0.0497		0.0006		0.0520		0.1465		0.0994	
	(0.2704)		(0.2784)		(0.2675)		(0.2569)		(0.2582)		(0.2570)		(0.2675)		(0.4167)	
d04MtlG	0.3052		0.3485		0.2389		0.1613		0.1262		0.1575		0.2389		0.2055	
	(0.2832)		(0.2867)		(0.2740)		(0.2765)		(0.2759)		(0.2759)		(0.2740)		(0.3867)	
d07Concordia	-0.6029		-0.6163		-0.8837		-0.8212		-0.8481		-0.8289		-0.8837		-0.8351	
	(0.8555)		(0.8482)		(0.8523)		(0.8515)		(0.8526)		(0.8501)		(0.8523)		(0.8511)	
d09UQAMG	-0.3552		-0.2677		-0.4876		-0.4625		-0.4919		-0.4582		-0.4876		-0.5465	
	(0.3259)		(0.3291)		(0.3251)		(0.3301)		(0.3321)		(0.3308)		(0.3251)		(0.5394)	
d08SherbrookeG	0.1165		0.1955		-0.0056		-0.0270		-0.0884		-0.0316		-0.0056		0.0024	
	(0.3867)		(0.3863)		(0.3770)		(0.3744)		(0.3764)		(0.3738)		(0.3770)		(0.4368)	
L_nbart_res			-0.1442	**												
			(0.0709)													
L1.ltotsubvmoyan3					0.1550	*	0.3077	*	0.3481	**	0.3059	*	0.1550	*	0.1625	
					(0.0910)		(0.1710)		(0.1674)		(0.1682)		(0.0910)		(0.1332)	
L1.sq_ltotsubvmoyan3					-0.0131	*	-0.0148	**	-0.0157	**	-0.0144	**	-0.0131	*	-0.0147	
					(0.0069)		(0.0074)		(0.0074)		(0.0072)		(0.0069)		(0.0141)	
L_subv_res							-0.1451									
							(0.1212)									
L_subv_res1									-0.1754							
									(0.1157)							
L_subv_res2											-0.1473					
											(0.1218)					
L_cont_res															-0.0285	
															(0.1623)	
_cons	-0.4640		-0.3533		-0.5218		-1.8202		-2.1817		-1.8444		-0.5218		-0.3867	
	(0.6131)		(0.6120)		(0.6613)		(1.4044)		(1.3490)		(1.4119)		(0.6613)		(0.6825)	
Inalpha																
_cons	4.1728	***	4.1607	***	4.1697	***	4.1589	***	4.1656	***	4.1589	***	4.1697	***	4.1604	***
	(0.0697)		(0.0696)		(0.0693)		(0.0696)		(0.0695)		(0.0696)		(0.0693)		(0.0694)	
Statistics																
N	8952		8771		8946		8769		8929		8769		8946		8769	
df_m	22		23		23		24		24		24		23		24	
k	24		25		25		26		26		26		25		26	
k_aux	1		1		1		1		1		1		1		1	
k_eq	2		2		2		2		2		2		2		2	
k_eq_model	1		1		1		1		1		1		1		1	
k_dv	1		1		1		1		1		1		1		1	
ll	-4388.47		-4338.23		-4387.04		-4337.43		-4385.1		-4337.43		-4387.04		-4338.02	
ll_0	-4476.75		-4424.43		-4476.38		-4424.31		-4475.32		-4424.31		-4476.38		-4424.31	
chi2	147.5116		151.7923		143.7985		145.2542		147.5943		144.6775		143.7985		142.7067	
aic	8824.947		8726.451		8824.075		8726.857		8822.201		8726.855		8824.075		8728.037	



Nbreg du Bloc B – breveter avec revendications								
Variable	nbreg_vce	nbreg_art	nbreg_vce1	nbreg_subv	nbreg_subv1	nbreg_subv2	nbreg_vce2	nbreg_cont
Endogénéité		x		x	x	x		x
bic	8995.338	8903.431	9001.549	8910.91	9006.724	8910.908	9001.549	8912.091
p	1.41E-20	5.88E-21	1.83E-19	2.51E-19	9.26E-20	3.21E-19	1.83E-19	7.40E-19

Tableau O-2: Régressions à binomiale négative en panel (deuxième étape) du bloc B – breveter avec revendications

xtnbreg du bloc B – breveter avec revendications																
Variable	xtnbreg_0		xtnbreg_art		xtnbreg_1		xtnbreg_subv		xtnbreg_subv1		xtnbreg_subv2		xtnbreg_2		xtnbreg_cont	
Endogénéité	x		x		x		x		x		x		x		x	
L1.nbarticlemoyan3	0.0753	***	0.1641	***												
	(0.0151)		(0.0477)													
L1.ltotcontmoyan3	0.0516	***	0.0471	***	0.0578	***	0.0335	***	0.0361	***	0.0344	***	0.0578	***	0.0982	***
	(0.0085)		(0.0086)		(0.0091)		(0.0105)		(0.0104)		(0.0105)		(0.0091)		(0.0332)	
L1.lmcliquind23	0.2187	**	0.1455		0.3036	***	0.3155	***	0.3138	***	0.3157	***	0.3036	***	0.2963	***
	(0.0874)		(0.0958)		(0.0816)		(0.0834)		(0.0828)		(0.0833)		(0.0816)		(0.0826)	
L1.sq_lmcliquind23	-0.0091	***	-0.0059		-0.0128	***	-0.0131	***	-0.0131	***	-0.0131	***	-0.0128	***	-0.0125	***
	(0.0035)		(0.0039)		(0.0032)		(0.0033)		(0.0033)		(0.0033)		(0.0032)		(0.0033)	
deg_app0	0.0962	***	0.0808	**	0.1213	***	0.1077	***	0.1089	***	0.1085	***	0.1213	***	0.1193	***
	(0.0312)		(0.0324)		(0.0306)		(0.0309)		(0.0307)		(0.0309)		(0.0306)		(0.0308)	
L1.avg_pat_past	0.4968	***	0.4710	***	0.5278	***	0.5064	***	0.5096	***	0.5064	***	0.5278	***	0.5028	***
	(0.0342)		(0.0362)		(0.0325)		(0.0326)		(0.0325)		(0.0325)		(0.0325)		(0.0378)	
nano_excl	-0.5282		-0.6762	*	-0.5823	*	-0.8178	**	-0.5867	*	-0.8161	**	-0.5823	*	-0.7795	**
	(0.3422)		(0.3873)		(0.3419)		(0.3860)		(0.3423)		(0.3860)		(0.3419)		(0.3856)	
bio_excl	-0.4877	***	-0.4040	***	-0.5608	***	-0.4782	***	-0.4880	***	-0.4810	***	-0.5608	***	-0.5498	***
	(0.1104)		(0.1170)		(0.1092)		(0.1109)		(0.1106)		(0.1109)		(0.1092)		(0.1096)	
d1997	-0.3315		-0.3613		-0.2188		0.2202		0.2085		0.2046		-0.2188		-0.2152	
	(0.2261)		(0.2264)		(0.2285)		(0.2517)		(0.2505)		(0.2518)		(0.2285)		(0.2288)	
d1998	-0.2209		-0.3236		-0.1569		-0.2701		-0.2173		-0.2673		-0.1569		-0.2396	
	(0.2159)		(0.2190)		(0.2158)		(0.2175)		(0.2157)		(0.2175)		(0.2158)		(0.2183)	
d1999	0.0895		-0.0324		0.1484		-0.0255		0.0255		-0.0215		0.1484		0.0694	
	(0.2069)		(0.2110)		(0.2067)		(0.2093)		(0.2082)		(0.2094)		(0.2067)		(0.2086)	
d2000	0.1553		0.0436		0.2012		0.0473		0.0851		0.0497		0.2012		0.1368	
	(0.2124)		(0.2165)		(0.2126)		(0.2154)		(0.2147)		(0.2155)		(0.2126)		(0.2134)	
d2001	-0.2011		-0.3139		-0.1252		-0.2856		-0.2632		-0.2815		-0.1252		-0.2053	
	(0.2266)		(0.2309)		(0.2260)		(0.2282)		(0.2281)		(0.2282)		(0.2260)		(0.2289)	
d2002	-0.0498		-0.1645		0.0327		-0.1877		-0.1573		-0.1820		0.0327		-0.0467	
	(0.2206)		(0.2249)		(0.2201)		(0.2245)		(0.2242)		(0.2246)		(0.2201)		(0.2227)	
d2003	-0.2294		-0.3313		-0.1600		-0.4093	*	-0.3781	*	-0.4026	*	-0.1600		-0.2221	
	(0.2215)		(0.2240)		(0.2219)		(0.2272)		(0.2270)		(0.2273)		(0.2219)		(0.2225)	
d2004	-0.6437	***	-0.7552	***	-0.5625	**	-0.8401	***	-0.8093	***	-0.8328	***	-0.5625	**	-0.6324	***
	(0.2378)		(0.2412)		(0.2393)		(0.2454)		(0.2451)		(0.2455)		(0.2393)		(0.2414)	
d2005	-1.1453	***	-1.2689	***	-1.0454	***	-1.3794	***	-1.3425	***	-1.3697	***	-1.0454	***	-1.1083	***
	(0.2674)		(0.2709)		(0.2686)		(0.2768)		(0.2764)		(0.2768)		(0.2686)		(0.2698)	
d01Laval	0.0820		0.1248		0.0202		-0.0795		-0.0823		-0.0761		0.0202		-0.0923	
	(0.1370)		(0.1392)		(0.1360)		(0.1379)		(0.1379)		(0.1378)		(0.1360)		(0.1600)	
d04MtlG	0.1849		0.2077		0.0970		-0.0156		0.0017		-0.0137		0.0970		0.0090	
	(0.1304)		(0.1323)		(0.1294)		(0.1314)		(0.1310)		(0.1315)		(0.1294)		(0.1395)	
d07Concordia	-0.8451	*	-0.7912	*	-0.8814	*	-0.7540		-0.7833	*	-0.7574		-0.8814	*	-0.8487	*

xtnbreg du bloc B – breveter avec revendications																
Variable	xtnbreg_0		xtnbreg_art		xtnbreg_1		xtnbreg_subv		xtnbreg_subv1		xtnbreg_subv2		xtnbreg_2		xtnbreg_cont	
Endogénéité	x				x				x				x			
d09UQAMG	(0.4602)		(0.4604)		(0.4607)		(0.4615)		(0.4614)		(0.4615)		(0.4607)		(0.4608)	
	-0.0331		0.0376		-0.1109		-0.1164		-0.1341		-0.1155		-0.1109		-0.2139	
	(0.1605)		(0.1640)		(0.1614)		(0.1616)		(0.1617)		(0.1615)		(0.1614)		(0.1808)	
d08SherbrookeG	0.0934		0.1603		0.0103		-0.0914		-0.1161		-0.0877		0.0103		-0.0438	
	(0.1915)		(0.1929)		(0.1910)		(0.1933)		(0.1940)		(0.1933)		(0.1910)		(0.1978)	
L_nbart_xt_ue			-0.1174	**												
			(0.0596)													
L1.ltotsubvmoyan3					-0.0072		0.1969	***	0.1895	***	0.1872	***	-0.0072		0.0224	
					(0.0488)		(0.0694)		(0.0684)		(0.0690)		(0.0488)		(0.0545)	
L1.sq_ltotsubvmoyan3					0.0008		-0.0014		-0.0013		-0.0012		0.0008		-0.0023	
					(0.0035)		(0.0035)		(0.0035)		(0.0035)		(0.0035)		(0.0043)	
L_subv_xt_ue							-0.1942	***								
							(0.0461)									
L_subv_xt_ue1									-0.1918	***						
									(0.0452)							
L_subv_x~ue2											-0.1867	***				
											(0.0460)					
L_cont_xt_ue																-0.0448
																(0.0335)
_cons	-5.1709	***	-5.1007	***	-5.2772	***	-6.7865	***	-6.7761	***	-6.7268	***	-5.2772	***	-5.1863	***
	(0.3292)		(0.3320)		(0.3665)		(0.5353)		(0.5242)		(0.5344)		(0.3665)		(0.3718)	
ln_r																
_cons	1.2887	***	1.3845	***	1.3761	***	1.5382	***	1.3953	***	1.5424	***	1.3761	***	1.4651	***
	(0.4080)		(0.4479)		(0.4478)		(0.5195)		(0.4550)		(0.5213)		(0.4478)		(0.4914)	
ln_s																
_cons	5.7593	***	5.8748	***	5.8757	***	6.0706	***	5.8976	***	6.0763	***	5.8757	***	5.9828	***
	(0.5334)		(0.5716)		(0.5761)		(0.6446)		(0.5826)		(0.6464)		(0.5761)		(0.6186)	
Statistics																
N	8952		8771		8946		8769		8929		8769		8946		8769	
N_g	1469		1460		1469		1460		1469		1460		1469		1460	
n_clust																
df_m	22		23		23		24		24		24		23		24	
k	25		26		26		27		27		27		26		27	
k_aux	2		2		2		2		2		2		2		2	
k_eq	3		3		3		3		3		3		3		3	
k_eq_model	1		1		1		1		1		1		1		1	
k_dv	1		1		1		1		1		1		1		1	
ll	-4280.36		-4231.68		-4290.54		-4234.89		-4280.32		-4235.55		-4290.54		-4243.16	
ll_0																
ll_c	-4285.04		-4235.56		-4294.39		-4237.77		-4284.08		-4238.41		-4294.39		-4246.35	
chi2	647.5107		623.5317		637.3693		642.0651		656.1224		642.0505		637.3693		602.697	
chi2_p																

xtnbreg du bloc B – breveter avec revendications								
Variable	xtnbreg_0	xtnbreg_art	xtnbreg_1	xtnbreg_subv	xtnbreg_subv1	xtnbreg_subv2	xtnbreg_2	xtnbreg_cont
Endogénéité		x		x	x	x		x
chi2_c	9.35946	7.762427	7.703792	5.75082	7.521327	5.709437	7.703792	6.372797
aic	8610.711	8515.362	8633.084	8523.787	8614.637	8525.107	8633.084	8540.32
bic	8788.201	8699.421	8817.657	8714.92	8806.257	8716.24	8817.657	8731.452
g_min	1	1	1	1	1	1	1	1
g_avg	6.093942	6.007534	6.089857	6.006164	6.078285	6.006164	6.089857	6.006164
g_max	10	10	10	10	10	10	10	10
r	3.628059	3.992888	3.959343	4.656236	4.036041	4.675607	3.959343	4.327773
s	317.1135	355.946	356.2909	432.9578	364.1466	435.3954	356.2909	396.5341
p	8.9E-123	5.3E-117	6.6E-120	3.7E-120	4.1E-123	3.7E-120	6.6E-120	6.5E-112

Tableau O-3: Régressions à binomiale négative à zéros-augmenté (deuxième étape) du bloc B  
– breveter avec revendications

zinb du bloc B – breveter avec revendications						
Variable	zinb_vce	zinb_art	zinb_subv	zinb_subv1	zinb_subv2	zinb_cont
Endogénéité		x	x	x	x	x
nbclaims0						
L_nbarticlemoyan3	0.0079 (0.0187)					
L_ltotsubvmoyan3	-0.0380 (0.0679)					-0.0854 (0.0651)
L_sq_ltotsubvmoyan3	0.0004 (0.0050)					0.0056 (0.0048)
L_ltotcontmoyan3	0.0068 (0.0118)	-0.0028 (0.0116)	-0.0074 (0.0163)	-0.0053 (0.0159)	-0.0125 (0.0178)	
L_lmcliqind23	0.2228 (0.1511)	0.0680 (0.1352)	0.2357 * (0.1431)	0.2322 (0.1459)	0.2341 * (0.1386)	0.2275 (0.1424)
L_sq_lmcliqind23	-0.0083 (0.0060)	-0.0020 (0.0053)	-0.0088 (0.0056)	-0.0087 (0.0058)	-0.0087 (0.0054)	-0.0085 (0.0056)
deg_app0	0.0168 (0.0372)	-0.0162 (0.0387)	0.0099 (0.0360)	0.0088 (0.0358)	0.0031 (0.0371)	0.0263 (0.0343)
nbloop0	0.2041 *** (0.0428)	0.2114 *** (0.0438)	0.2062 *** (0.0466)	0.2135 *** (0.0486)	0.2068 *** (0.0456)	0.2137 *** (0.0454)
propuninv0	-0.3649 (0.2704)	-0.2834 (0.3122)	-0.3199 (0.3360)	-0.2647 (0.3515)	-0.3155 (0.3290)	-0.0903 (0.3634)
nano_excl	0.2484 (0.4730)	0.0873 (0.4886)	0.1136 (0.5034)	0.3787 (0.5191)	0.1370 (0.4934)	0.1804 (0.5228)
bio_excl	-0.2210 (0.1638)	-0.1152 (0.1504)	-0.1875 (0.1508)	-0.1932 (0.1512)	-0.1749 (0.1500)	-0.1868 (0.1565)
d1997	-0.1022 (0.2572)	-0.0775 (0.2675)	0.1602 (0.3512)	0.1387 (0.3477)	0.2612 (0.3711)	-0.0398 (0.2718)
d1998	0.1315 (0.2814)	0.0296 (0.2756)	0.1085 (0.2778)	0.0942 (0.2819)	0.0999 (0.2721)	0.1622 (0.2789)
d1999	0.0115 (0.2409)	-0.1718 (0.2239)	-0.0524 (0.2280)	-0.0495 (0.2321)	-0.0794 (0.2215)	0.0664 (0.2328)
d2000	0.2092 (0.2613)	-0.0446 (0.2455)	0.1053 (0.2434)	0.1509 (0.2485)	0.0774 (0.2378)	0.2501 (0.2571)
d2001	0.0313 (0.2674)	-0.1741 (0.2549)	-0.0352 (0.2550)	-0.0375 (0.2577)	-0.0695 (0.2512)	0.1088 (0.2599)
d2002	0.3531 (0.2938)	0.2093 (0.2871)	0.2641 (0.2887)	0.2841 (0.2906)	0.2345 (0.2863)	0.4411 (0.2912)
d2003	0.0831 (0.2695)	-0.0724 (0.2630)	-0.0169 (0.2680)	-0.0060 (0.2702)	-0.0518 (0.2658)	0.2292 (0.3016)
d2004	-0.2172 (0.2718)	-0.3815 (0.2648)	-0.3152 (0.2664)	-0.2957 (0.2705)	-0.3494 (0.2651)	-0.2129 (0.2717)
d2005	0.1542 (0.3483)	-0.1027 (0.3593)	-0.0063 (0.3712)	0.0217 (0.3732)	-0.0515 (0.3758)	0.1875 (0.3686)
d01Laval	0.0144 (0.1734)	0.0695 (0.1715)	-0.0734 (0.1672)	-0.0724 (0.1668)	-0.0945 (0.1708)	0.1405 (0.1734)
d04MtIG	0.1035 (0.1867)	0.1562 (0.1819)	0.0403 (0.1772)	0.0604 (0.1786)	0.0047 (0.1763)	0.1746 (0.1805)
d07Concordia	0.7133 ** (0.3184)	0.7371 ** (0.3068)	0.7296 ** (0.3205)	0.7239 ** (0.3205)	0.7265 ** (0.3193)	0.7030 ** (0.3240)
d09UQAMG	0.2639 (0.3042)	0.3561 (0.2902)	0.2215 (0.2857)	0.2172 (0.2855)	0.2168 (0.2832)	0.4319 (0.3122)
d08SherbrookeG	0.0075 (0.2588)	0.1357 (0.2464)	-0.0219 (0.2498)	-0.0205 (0.2512)	-0.0285 (0.2517)	0.0709 (0.2520)
L_nbart_pre		0.0864 ** (0.0366)				
L_subv_pre			0.0595 (0.0712)			
L_subv1_pre				0.0449 (0.0671)		
L_subv2_pre					0.0967 (0.0830)	
L_cont_pre						-0.0412 ** (0.0187)
_cons	2.9707 *** (0.5202)	2.8822 *** (0.4553)	2.1066 ** (0.8750)	2.2162 *** (0.8386)	1.7748 * (0.9749)	2.7204 *** (0.5619)

zinb du bloc B – breveter avec revendications						
Variable	zinb_vce	zinb_art	zinb_subv	zinb_subv1	zinb_subv2	zinb_cont
Endogénéité		x	x	x	x	x
inflate						
L_nbarticlemoyan3	-0.0449 ** (0.0211)					
L_ltotsubvmoyan3	-0.0930 * (0.0525)					-0.0204 (0.0558)
L_sq_ltotsubvmoyan3	0.0081 ** (0.0038)					-0.0004 (0.0040)
L_ltotcontmoyan3	-0.0550 *** (0.0110)	-0.0529 *** (0.0113)	-0.0529 *** (0.0113)	-0.0546 *** (0.0113)	-0.0529 *** (0.0113)	
L_lmcliqind23	-0.3744 *** (0.0845)	-0.3406 *** (0.0850)	-0.3379 *** (0.0850)	-0.3413 *** (0.0846)	-0.3380 *** (0.0850)	-0.3413 *** (0.0851)
L_sq_lmcliqind23	0.0149 *** (0.0034)	0.0142 *** (0.0034)	0.0141 *** (0.0034)	0.0143 *** (0.0034)	0.0141 *** (0.0034)	0.0143 *** (0.0034)
deg_app0	-0.0818 ** (0.0362)	-0.1230 *** (0.0371)	-0.1228 *** (0.0371)	-0.1239 *** (0.0368)	-0.1229 *** (0.0371)	-0.1147 *** (0.0365)
L_o_patpast	-1.0123 *** (0.0715)					
nano_excl	0.1948 (0.4042)	0.9579 * (0.4943)	0.9597 * (0.4938)	0.7793 * (0.4024)	0.9605 * (0.4934)	1.0276 ** (0.4868)
bio_excl	0.2602 * (0.1471)	0.4003 *** (0.1534)	0.4005 *** (0.1533)	0.4065 *** (0.1519)	0.4006 *** (0.1534)	0.3988 *** (0.1537)
L_avg_pat_past		-2.5042 *** (0.3369)	-2.5133 *** (0.3371)	-2.5905 *** (0.3536)	-2.5094 *** (0.3353)	-2.5812 *** (0.3656)
_cons	4.6389 *** (0.3169)	4.0972 *** (0.2479)	4.0923 *** (0.2479)	4.1226 *** (0.2471)	4.0922 *** (0.2480)	4.1097 *** (0.3256)
Inalpha						
_cons	-0.1041 (0.0913)	-0.0352 (0.1236)	-0.0228 (0.1209)	-0.0077 (0.1257)	-0.0254 (0.1212)	-0.0170 (0.1369)
Statistics						
N	8946	8771	8775	8935	8775	8772
df_m	25	23	23	23	23	24
k	38	33	33	33	33	35
k_aux	1	1	1	1	1	1
k_eq	3	3	3	3	3	3
k_eq_model	1	1	1	1	1	1
k_dv	1	1	1	1	1	1
ll	-4030.23	-4044.17	-4048.5	-4088.44	-4047.74	-4059.07
ll_0	-4075.21	-4084.11	-4084.3	-4128.21	-4084.3	-4099.58
chi2	73.59485	74.33424	61.55077	60.87431	64.22707	69.74233
aic	8136.467	8154.332	8163.007	8242.878	8161.476	8188.147
bic	8406.227	8387.946	8396.636	8477.103	8395.105	8435.923
p	1.11E-06	2.56E-07	2.27E-05	2.85E-05	9.15E-06	2.39E-06

Tableau O-4: Régressions tobit (deuxième étape) du bloc B – breveter avec revendications

tobit du bloc B – breveter avec revendications					
Variable	HSM_tobit_art	HSM_tobit_subv	HSM_tobit_subv_1	HSM_tobit_subv_2	HSM_tobit_cont
Endogénéité	x	x	x	x	x
model					
L_ltotcontmoyan3	0.1189 *** (0.0310)	0.0445 (0.0401)	0.0520 (0.0401)	0.0487 (0.0400)	0.1303 (0.2168)
L_lmcliqind23	0.4043 (0.2990)	0.8748 *** (0.2805)	0.8644 *** (0.2779)	0.8700 *** (0.2791)	0.7172 *** (0.2777)
L_sq_lmcliqind23	-0.0166 (0.0120)	-0.0358 *** (0.0111)	-0.0354 *** (0.0110)	-0.0356 *** (0.0111)	-0.0302 *** (0.0110)
deg_app0	0.1958 ** (0.0939)	0.2306 ** (0.0896)	0.2363 *** (0.0885)	0.2345 *** (0.0898)	0.3017 *** (0.0890)
L_avg_pat_past	2.4340 *** (0.7288)	2.5230 *** (0.7271)	2.5830 *** (0.7440)	2.5233 *** (0.7305)	2.6788 *** (0.7387)
nano_excl	-2.0066 ** (1.0125)	-2.1228 ** (1.0142)	-1.6122 * (0.9450)	-2.1367 ** (1.0118)	-2.2459 ** (1.0080)
bio_excl	-1.0406 *** (0.3940)	-1.1019 *** (0.3886)	-1.1298 *** (0.3870)	-1.1129 *** (0.3884)	-1.2862 *** (0.3852)
L_nbarticlemoyan3	0.2901 *** (0.1065)				
L_nbart_res	-0.1157 (0.1279)				
d1997	-0.8626 (0.5565)	1.0274 (0.7535)	0.9719 (0.7412)	0.9252 (0.7510)	-0.7055 (0.5645)
d1998	-0.9315 * (0.5305)	-0.7711 (0.5343)	-0.6308 (0.5290)	-0.7726 (0.5339)	-0.8020 (0.5337)
d1999	-0.1716 (0.5421)	-0.1746 (0.5399)	-0.0259 (0.5373)	-0.1614 (0.5407)	0.0597 (0.5459)
d2000	0.1838 (0.5787)	0.3007 (0.5745)	0.4156 (0.5698)	0.3077 (0.5745)	0.4788 (0.5909)
d2001	-0.7904 (0.6179)	-0.7059 (0.6166)	-0.6463 (0.6148)	-0.6978 (0.6167)	-0.4857 (0.6233)
d2002	-0.4148 (0.6084)	-0.4680 (0.6054)	-0.3764 (0.6036)	-0.4497 (0.6060)	-0.1183 (0.6012)
d2003	-0.8867 (0.6148)	-1.0090 (0.6191)	-0.9243 (0.6184)	-0.9967 (0.6190)	-0.6622 (0.6115)
d2004	-1.9095 *** (0.6418)	-2.0508 *** (0.6509)	-1.9664 *** (0.6496)	-2.0370 *** (0.6514)	-1.6459 *** (0.6390)
d2005	-3.1359 *** (0.6637)	-3.2959 *** (0.6796)	-3.1953 *** (0.6803)	-3.2751 *** (0.6807)	-2.7962 *** (0.6683)
d01Laval	0.3495 (0.4995)	-0.2205 (0.4952)	-0.2344 (0.4959)	-0.1996 (0.4965)	0.1134 (0.6749)
d04MtIG	0.4450 (0.4730)	-0.1181 (0.4560)	-0.0724 (0.4552)	-0.1058 (0.4560)	0.2048 (0.6033)
d07Concordia	-1.7142 (1.8688)	-1.6374 (1.8733)	-1.6835 (1.8659)	-1.6560 (1.8731)	-1.9323 (1.8634)
d09UQAMG	-0.1249 (0.5805)	-0.3796 (0.5860)	-0.4311 (0.5845)	-0.3761 (0.5863)	-0.4235 (0.8291)
d08SherbrookeG	0.3542 (0.7281)	-0.2139 (0.7436)	-0.2788 (0.7463)	-0.2012 (0.7428)	0.0740 (0.7969)
L_ltotsubvmoyan3		0.7774 *** (0.2668)	0.7543 *** (0.2628)	0.7210 *** (0.2657)	0.0262 (0.1832)
L_sq_ltotsubvmoyan3		-0.0116 (0.0123)	-0.0113 (0.0124)	-0.0099 (0.0123)	-0.0015 (0.0185)
L_subv_res		-0.6789 *** (0.1915)			
L_subv_res1			-0.6646 *** (0.1867)		
L_subv_res2				-0.6401 *** (0.1919)	

tobit du bloc B – breveter avec revendications					
Variable	HSM_tobit_art	HSM_tobit_subv	HSM_tobit_subv_1	HSM_tobit_subv_2	HSM_tobit_cont
Endogénéité	x	x	x	x	x
L_cont_res					0.0016 (0.2184)
_cons	-10.1429 *** (1.0460)	-16.8372 *** (2.0976)	-16.7602 *** (2.0420)	-16.4784 *** (2.0876)	-10.5250 *** (1.1553)
sigma _cons	5.4205 *** (0.1127)	5.4240 *** (0.1135)	5.4302 *** (0.1134)	5.4268 *** (0.1132)	5.4428 *** (0.1112)
Statistics					
N	8771	8769	8929	8769	8769
df_m	23	24	24	24	24
k_aux	1	1	1	1	1
ll	-2845.33	-2845.08	-2873.96	-2846.34	-2854.61
ll_0	-3039.04	-3038.91	-3073.51	-3038.91	-3038.91
aic	5740.661	5742.165	5799.929	5744.677	5761.216
bic	5917.642	5926.219	5984.452	5928.731	5945.269
r2_p	0.06374	0.063783	0.064925	0.06337	0.060649
p	1.26E-20	5.82E-21	5.83E-22	3.48E-20	7.03E-17



## ANNEXE P – Régressions de deuxième étape du bloc C

Tableau P-1: Régressions à binomiale négative (deuxième étape) du bloc C – breveter avec citations

Nbreg du Bloc C – breveter avec citations										
Variable	nbreg_vce	nbreg_art	nbreg_vce1	nbreg_subv	nbreg_subv1	nbreg_subv2	nbreg_vce2	nbreg_cont		
Endogénéité	x		x		x		x		x	
L1.nbarticlemoyan3	0.0016 (0.0359)	0.0085 (0.0658)								
L1.ltotcontmoyan3	0.0390 * (0.0210)	0.0362 * (0.0216)	0.0485 ** (0.0219)	0.0347 (0.0299)	0.0353 (0.0278)	0.0357 (0.0306)	0.0485 ** (0.0219)	-0.0958 (0.1725)		
L1.lmcliquind23	0.2934 (0.2318)	0.2926 (0.2404)	0.2829 (0.2238)	0.2988 (0.2263)	0.2962 (0.2229)	0.3002 (0.2265)	0.2829 (0.2238)	0.2932 (0.2254)		
L1.sq_lmcliquind23	-0.0119 (0.0092)	-0.0119 (0.0096)	-0.0114 (0.0089)	-0.0119 (0.0090)	-0.0118 (0.0088)	-0.0120 (0.0090)	-0.0114 (0.0089)	-0.0119 (0.0089)		
deg_app0	0.1704 ** (0.0728)	0.1660 ** (0.0765)	0.1794 ** (0.0723)	0.1687 ** (0.0759)	0.1700 ** (0.0749)	0.1698 ** (0.0760)	0.1794 ** (0.0723)	0.1664 ** (0.0739)		
L1.avg_pat_past	4.6925 *** (0.9772)	4.6721 *** (0.9715)	4.7922 *** (0.9803)	4.7362 *** (0.9825)	4.7393 *** (0.9831)	4.7398 *** (0.9816)	4.7922 *** (0.9803)	4.8455 *** (0.9859)		
nano_excl	-1.0236 * (0.5669)	-1.1004 * (0.6223)	-1.0326 * (0.5690)	-1.0362 (0.6409)	-0.9555 * (0.5763)	-1.0462 (0.6396)	-1.0326 * (0.5690)	-1.2291 * (0.6489)		
bio_excl	-1.3143 *** (0.2740)	-1.2808 *** (0.2801)	-1.3478 *** (0.2639)	-1.3143 *** (0.2675)	-1.3430 *** (0.2659)	-1.3149 *** (0.2675)	-1.3478 *** (0.2639)	-1.2872 *** (0.2627)		
d1997	-0.1824 (0.4354)	-0.1796 (0.4313)	-0.2763 (0.4625)	-0.0570 (0.6029)	-0.0110 (0.5817)	-0.0791 (0.6107)	-0.2763 (0.4625)	-0.2710 (0.4630)		
d1998	-0.8664 ** (0.3473)	-0.9333 *** (0.3530)	-0.8695 ** (0.3512)	-0.9070 ** (0.3559)	-0.8370 ** (0.3511)	-0.9091 ** (0.3551)	-0.8695 ** (0.3512)	-0.8544 ** (0.3812)		
d1999	-0.9358 *** (0.3620)	-0.9870 *** (0.3644)	-0.8913 ** (0.3571)	-0.9724 *** (0.3634)	-0.9269 ** (0.3608)	-0.9654 *** (0.3625)	-0.8913 ** (0.3571)	-0.8704 ** (0.3717)		
d2000	-0.6766 * (0.3839)	-0.7014 * (0.3863)	-0.6404 * (0.3783)	-0.6707 * (0.3809)	-0.6478 * (0.3792)	-0.6681 * (0.3800)	-0.6404 * (0.3783)	-0.5747 (0.4093)		
d2001	-0.8464 * (0.4409)	-0.8772 ** (0.4395)	-0.8491 ** (0.4165)	-0.9201 ** (0.4148)	-0.9019 ** (0.4137)	-0.9132 ** (0.4138)	-0.8491 ** (0.4165)	-0.7801 * (0.4359)		
d2002	-0.7712 * (0.4450)	-0.8191 * (0.4451)	-0.7324 * (0.4306)	-0.8039 * (0.4313)	-0.7630 * (0.4310)	-0.8003 * (0.4306)	-0.7324 * (0.4306)	-0.7408 * (0.4391)		
d2003	-2.0406 *** (0.4472)	-2.0799 *** (0.4459)	-2.0113 *** (0.4440)	-2.0838 *** (0.4507)	-2.0513 *** (0.4503)	-2.0779 *** (0.4487)	-2.0113 *** (0.4440)	-2.0635 *** (0.4464)		
d2004	-3.8225 *** (0.6345)	-3.8593 *** (0.6303)	-3.7520 *** (0.6375)	-3.8329 *** (0.6293)	-3.8084 *** (0.6311)	-3.8288 *** (0.6297)	-3.7520 *** (0.6375)	-3.8042 *** (0.6352)		
d2005	-3.6370 *** (0.6387)	-3.6856 *** (0.6355)	-3.5808 *** (0.6361)	-3.6759 *** (0.6409)	-3.6442 *** (0.6408)	-3.6695 *** (0.6406)	-3.5808 *** (0.6361)	-3.6756 *** (0.6341)		
d01Laval	0.2031 (0.3996)	0.2157 (0.4039)	0.2221 (0.3950)	0.1563 (0.3621)	0.1360 (0.3665)	0.1661 (0.3617)	0.2221 (0.3950)	0.5740 (0.4836)		

Nbreg du Bloc C – breveter avec citations																
Variable	nbreg_vce		nbreg_art		nbreg_vce1		nbreg_subv		nbreg_subv1		nbreg_subv2		nbreg_vce2		nbreg_cont	
Endogénéité			x				x		x		x				x	
d04MtIG	0.5726		0.5808		0.6136		0.5409		0.5328		0.5501		0.6136		0.8941	**
	(0.3863)		(0.3929)		(0.3806)		(0.3393)		(0.3442)		(0.3377)		(0.3806)		(0.4241)	
d07Concordia	-1.1359		-1.0950		-1.1726		-1.1306		-1.1538		-1.1325		-1.1726		-1.1167	
	(0.7572)		(0.7492)		(0.7581)		(0.7519)		(0.7558)		(0.7484)		(0.7581)		(0.7478)	
d09UQAMG	0.0109		0.0261		0.0300		-0.0038		-0.0222		0.0035		0.0300		0.4666	
	(0.4527)		(0.4564)		(0.4497)		(0.4204)		(0.4233)		(0.4220)		(0.4497)		(0.5796)	
d08SherbrookeG	0.4730		0.4975		0.4823		0.4480		0.4217		0.4533		0.4823		0.6967	
	(0.5126)		(0.5070)		(0.5059)		(0.4947)		(0.4977)		(0.4940)		(0.5059)		(0.5225)	
L_nbart_res			-0.0158													
			(0.0837)													
L1.ltotsubvmoyan3					0.0253		0.1133		0.1315		0.1012		0.0253		-0.0716	
					(0.1014)		(0.1851)		(0.1731)		(0.1842)		(0.1014)		(0.1485)	
L1.sq_ltotsubvmoyan3					-0.0049		-0.0060		-0.0065		-0.0057		-0.0049		0.0067	
					(0.0077)		(0.0078)		(0.0078)		(0.0077)		(0.0077)		(0.0155)	
L_subv_res							-0.0831									
							(0.1434)									
L_subv_res1									-0.0963							
									(0.1291)							
L_subv_res2											-0.0739					
											(0.1454)					
L_cont_res															0.1475	
															(0.1778)	
_cons	-2.9627	***	-2.9216	***	-2.7169	***	-3.3753	**	-3.5392	**	-3.3043	*	-2.7169	***	-2.7865	***
	(0.7880)		(0.7883)		(0.8346)		(1.6682)		(1.5164)		(1.6994)		(0.8346)		(0.8409)	
Inalpha																
_cons	3.7435	***	3.7473	***	3.7419	***	3.7443	***	3.7378	***	3.7446	***	3.7419	***	3.7427	***
	(0.1107)		(0.1100)		(0.1101)		(0.1101)		(0.1107)		(0.1099)		(0.1101)		(0.1090)	
Statistics																
N	8952		8771		8946		8769		8929		8769		8946		8769	
df_m	22		23		23		24		24		24		23		24	
k	24		25		25		26		26		26		25		26	
k_aux	1		1		1		1		1		1		1		1	
k_eq	2		2		2		2		2		2		2		2	
k_eq_model	1		1		1		1		1		1		1		1	
k_dv	1		1		1		1		1		1		1		1	
ll	-1623.92		-1598.37		-1622.9		-1597.24		-1621.9		-1597.28		-1622.9		-1597.11	
ll_0	-1756.67		-1728		-1756.49		-1727.94		-1755.99		-1727.94		-1756.49		-1727.94	
chi2	160.9193		169.5347		175.1647		178.6108		179.633		178.2399		175.1647		177.9011	
aic	3295.832		3246.741		3295.806		3246.473		3295.806		3246.563		3295.806		3246.215	
bic	3466.223		3423.722		3473.28		3430.526		3480.33		3430.616		3473.28		3430.268	
p	4.07E-23		2.59E-24		2.18E-25		1.35E-25		8.61E-26		1.59E-25		2.18E-25		1.84E-25	

Tableau P-2: Régressions à binomiale négative en panel (deuxième étape) du bloc C – breveter avec citations

xtnbreg du bloc C – breveter avec citations																
Variable	xtnbreg_0		xtnbreg_art		xtnbreg_1		xtnbreg_subv		xtnbreg_subv1		xtnbreg_subv2		xtnbreg_2		xtnbreg_cont	
Endogénéité	X		X		X		X		X		X		x			
L1.nbarticlemoyan3	0.0601	**	0.1298	*												
	(0.0256)		(0.0770)													
L1.ltotcontmoyan3	0.0386	***	0.0327	**	0.0463	***	0.0294		0.0350	*	0.0308	*	0.0463	***	-0.0095	
	(0.0146)		(0.0148)		(0.0157)		(0.0186)		(0.0183)		(0.0186)		(0.0157)		(0.0779)	
L1.lmcliquind23	0.2172		0.1720		0.2896	**	0.2949	**	0.2888	**	0.2952	**	0.2896	**	0.2963	**
	(0.1491)		(0.1590)		(0.1434)		(0.1446)		(0.1434)		(0.1446)		(0.1434)		(0.1447)	
L1.sq_lmcliquind23	-0.0088		-0.0068		-0.0119	**	-0.0120	**	-0.0118	**	-0.0120	**	-0.0119	**	-0.0121	**
	(0.0059)		(0.0064)		(0.0057)		(0.0057)		(0.0057)		(0.0057)		(0.0057)		(0.0057)	
deg_app0	0.0820	*	0.0713		0.1048	**	0.0994	**	0.0996	**	0.1000	**	0.1048	**	0.1067	**
	(0.0483)		(0.0501)		(0.0478)		(0.0484)		(0.0479)		(0.0484)		(0.0478)		(0.0484)	
L1.avg_pat_past	0.2463	***	0.2290	***	0.2547	***	0.2431	***	0.2444	***	0.2437	***	0.2547	***	0.2714	***
	(0.0648)		(0.0679)		(0.0650)		(0.0659)		(0.0656)		(0.0659)		(0.0650)		(0.0690)	
nano_excl	-0.0096		-0.2646		-0.0641		-0.3411		-0.0211		-0.3444		-0.0641		-0.3891	
	(0.4990)		(0.5575)		(0.5038)		(0.5613)		(0.5062)		(0.5611)		(0.5038)		(0.5607)	
bio_excl	-0.7641	***	-0.6821	***	-0.8381	***	-0.7742	***	-0.8033	***	-0.7783	***	-0.8381	***	-0.8104	***
	(0.2118)		(0.2195)		(0.2128)		(0.2149)		(0.2149)		(0.2149)		(0.2128)		(0.2131)	
d1997	-0.5350	*	-0.5675	**	-0.4763	*	-0.2883		-0.2849		-0.3116		-0.4763	*	-0.4695	
	(0.2810)		(0.2822)		(0.2858)		(0.3323)		(0.3292)		(0.3324)		(0.2858)		(0.2868)	
d1998	-0.2525		-0.3551		-0.1802		-0.2895		-0.2181		-0.2849		-0.1802		-0.2217	
	(0.2594)		(0.2643)		(0.2590)		(0.2631)		(0.2601)		(0.2630)		(0.2590)		(0.2675)	
d1999	-0.1685		-0.2582		-0.1039		-0.2189		-0.1722		-0.2114		-0.1039		-0.1114	
	(0.2610)		(0.2667)		(0.2614)		(0.2688)		(0.2669)		(0.2690)		(0.2614)		(0.2705)	
d2000	-0.1411		-0.2370		-0.0236		-0.1343		-0.0857		-0.1270		-0.0236		-0.0144	
	(0.2844)		(0.2919)		(0.2851)		(0.2914)		(0.2891)		(0.2915)		(0.2851)		(0.3017)	
d2001	-0.6387	**	-0.7177	**	-0.5029		-0.6089	*	-0.5811	*	-0.6000	*	-0.5029		-0.4708	
	(0.3103)		(0.3166)		(0.3108)		(0.3173)		(0.3167)		(0.3174)		(0.3108)		(0.3274)	
d2002	-0.6596	**	-0.7377	**	-0.5099		-0.6503	**	-0.6167	*	-0.6376	**	-0.5099		-0.4841	
	(0.3096)		(0.3151)		(0.3104)		(0.3233)		(0.3221)		(0.3232)		(0.3104)		(0.3259)	
d2003	-1.3234	***	-1.3957	***	-1.1918	***	-1.3388	***	-1.3061	***	-1.3257	***	-1.1918	***	-1.1757	***
	(0.3434)		(0.3478)		(0.3478)		(0.3604)		(0.3596)		(0.3605)		(0.3478)		(0.3578)	
d2004	-2.5222	***	-2.6029	***	-2.3826	***	-2.5521	***	-2.5164	***	-2.5367	***	-2.3826	***	-2.3798	***
	(0.5070)		(0.5105)		(0.5096)		(0.5217)		(0.5207)		(0.5217)		(0.5096)		(0.5143)	

xtnbreg du bloc C – breveter avec citations																
Variable	xtnbreg_0		xtnbreg_art		xtnbreg_1		xtnbreg_subv		xtnbreg_subv1		xtnbreg_subv2		xtnbreg_2		xtnbreg_cont	
Endogénéité			X				X		X				X		x	
d2005	-2.3806	***	-2.4707	***	-2.2183	***	-2.4159	***	-2.3740	***	-2.3979	***	-2.2183	***	-2.2343	***
	(0.4747)		(0.4791)		(0.4778)		(0.4954)		(0.4942)		(0.4955)		(0.4778)		(0.4806)	
d01Laval	0.0681		0.1039		0.0137		-0.0377		-0.0314		-0.0332		0.0137		0.1187	
	(0.2639)		(0.2682)		(0.2654)		(0.2678)		(0.2678)		(0.2678)		(0.2654)		(0.3180)	
d04MtlG	0.1011		0.1306		0.0158		-0.0334		-0.0207		-0.0299		0.0158		0.0765	
	(0.2553)		(0.2592)		(0.2557)		(0.2577)		(0.2572)		(0.2578)		(0.2557)		(0.2813)	
d07Concordia	-0.6750		-0.6191		-0.7926		-0.7020		-0.7349		-0.7081		-0.7926		-0.7675	
	(0.6829)		(0.6840)		(0.6892)		(0.6927)		(0.6914)		(0.6926)		(0.6892)		(0.6917)	
d09UQAMG	-0.0147		0.0722		-0.0765		-0.0566		-0.0952		-0.0548		-0.0765		0.0711	
	(0.2950)		(0.3015)		(0.2982)		(0.2986)		(0.2986)		(0.2986)		(0.2982)		(0.3474)	
d08SherbrookeG	0.0421		0.1232		-0.0624		-0.0909		-0.1201		-0.0852		-0.0624		0.0325	
	(0.3705)		(0.3748)		(0.3730)		(0.3766)		(0.3777)		(0.3766)		(0.3730)		(0.3877)	
L_nbart_xt_ue			-0.0855													
			(0.0919)													
L1.ltotsubvmoyan3					0.0181		0.1126		0.1136		0.0996		0.0181		-0.0076	
					(0.0781)		(0.1177)		(0.1145)		(0.1170)		(0.0781)		(0.0848)	
L1.sq_ltotsubvmoyan3					-0.0028		-0.0034		-0.0039		-0.0031		-0.0028		0.0006	
					(0.0060)		(0.0061)		(0.0061)		(0.0061)		(0.0060)		(0.0072)	
L_subv_xt_ue							-0.0905									
							(0.0791)									
L_subv_xt_ue1									-0.0914							
									(0.0768)							
L_subv_xt_ue2											-0.0795					
											(0.0791)					
L_cont_xt_ue															0.0524	
															(0.0796)	
_cons	-2.9927	***	-2.9909	***	-2.9000	***	-3.6196	***	-3.6013	***	-3.5307	***	-2.9000	***	-2.9227	***
	(0.5562)		(0.5617)		(0.6015)		(0.8832)		(0.8565)		(0.8824)		(0.6015)		(0.6121)	
ln_r																
_cons	-0.2172	**	-0.2164	**	-0.2271	**	-0.2309	**	-0.2280	**	-0.2308	**	-0.2271	**	-0.2300	**
	(0.1014)		(0.1023)		(0.1006)		(0.1016)		(0.1009)		(0.1016)		(0.1006)		(0.1013)	
ln_s																
_cons	-1.3639	***	-1.3261	***	-1.4543	***	-1.4229	***	-1.4427	***	-1.4245	***	-1.4543	***	-1.4556	***
	(0.2090)		(0.2127)		(0.2002)		(0.2037)		(0.2007)		(0.2036)		(0.2002)		(0.2034)	
Statistics																
N	8952		8771		8946		8769		8929		8769		8946		8769	
N_g	1469		1460		1469		1460		1469		1460		1469		1460	

xtnbreg du bloc C – breveter avec citations								
Variable	xtnbreg_0	xtnbreg_art	xtnbreg_1	xtnbreg_subv	xtnbreg_subv1	xtnbreg_subv2	xtnbreg_2	xtnbreg_cont
Endogénéité		X		X	X	X		x
df_m	22	23	23	24	24	24	23	24
k	25	26	26	27	27	27	26	27
k_aux	2	2	2	2	2	2	2	2
k_eq	3	3	3	3	3	3	3	3
k_eq_model	1	1	1	1	1	1	1	1
k_dv	1	1	1	1	1	1	1	1
rank_0								
ll	-1554.12	-1532.62	-1556.19	-1534.92	-1555.07	-1535.07	-1556.19	-1535.35
ll_0								
ll_c	-1615.89	-1591.41	-1621.61	-1596.59	-1619.7	-1596.83	-1621.61	-1597.84
chi2	138.6169	136.9806	133.8207	132.4245	136.0165	132.0885	133.8207	132.5685
chi2_p								
chi2_c	123.5579	117.5928	130.8525	123.3409	129.2481	123.5063	130.8525	124.9788
aic	3158.232	3117.236	3164.375	3123.843	3164.145	3124.145	3164.375	3124.697
bic	3335.723	3301.295	3348.948	3314.975	3355.765	3315.277	3348.948	3315.83
g_min	1	1	1	1	1	1	1	1
g_avg	6.093942	6.007534	6.089857	6.006164	6.078285	6.006164	6.089857	6.006164
g_max	10	10	10	10	10	10	10	10
r	0.804764	0.805399	0.796813	0.793781	0.796094	0.793891	0.796813	0.794528
s	0.255667	0.265498	0.233561	0.241025	0.236288	0.240631	0.233561	0.23634
p	6.52E-19	3.35E-18	1.28E-17	5.64E-17	1.25E-17	6.49E-17	1.28E-17	5.31E-17

Tableau P-3: Régressions à binomiale négative à zéro-augmenté (deuxième étape) du bloc C  
– breveter avec citations

zinb du bloc C – breveter avec citations						
Variable	zinb_vce	zinb_art	zinb_subv	zinb_subv1	zinb_subv2	zinb_cont
Endogénéité	x	x	x	x	x	x
nbcitp50						
L_nbarticlemoyan3	-0.0460 ** (0.0395)					
L_ltotsubvmoyan3	-0.1893 (0.1831)					-0.6151 (0.1938)
L_sq_ltotsubvmoyan3	0.0066 ** (0.0126)					0.0456 ** (0.0146)
L_ltotcontmoyan3	0.0301 ** (0.0250)	0.0266 ** (0.0287)	0.0099 ** (0.0491)	0.0215 ** (0.0437)	-0.0030 ** (0.0498)	
L_lmcliqind23	0.7942 (0.3566)	0.4401 (0.2993)	0.7793 (0.3986)	0.7840 (0.3926)	0.7700 (0.3856)	0.2532 (0.2247)
L_sq_lmcliqind23	-0.0330 ** (0.0143)	-0.0179 ** (0.0119)	-0.0321 ** (0.0158)	-0.0323 ** (0.0156)	-0.0316 ** (0.0153)	-0.0109 *** (0.0090)
deg_app0	0.0784 * (0.0815)	-0.0727 * (0.0756)	-0.0156 * (0.0781)	-0.0094 * (0.0751)	-0.0298 * (0.0773)	-0.0200 * (0.0744)
nbloop0	0.4657 (0.1576)	0.3833 (0.1120)	0.4186 (0.1398)	0.4251 (0.1384)	0.4189 (0.1362)	0.6217 (0.1833)
propuninv0	10.5144 (0.6223)	10.0319 (0.6820)	9.8217 (0.6820)	9.7821 (0.6733)	9.8488 (0.6838)	9.7720 (0.6892)
nano_excl	-0.6134 (0.6145)	-0.8369 (0.4764)	-0.8530 (0.9578)	-0.9469 (0.4949)	-0.7962 (0.9020)	-0.4951 (0.4874)
bio_excl	-0.5317 (0.3612)	-0.3789 (0.3706)	-0.5459 (0.3715)	-0.5952 (0.3695)	-0.5180 (0.3706)	-0.3959 (0.3408)
d1997	1.0155 (0.5349)	0.6031 (0.7481)	0.8496 (1.0085)	0.7713 (0.9700)	1.0318 (1.0273)	0.6521 (0.5383)
d1998	0.2443 (0.5207)	-0.4972 (0.8207)	-0.4056 (0.7726)	-0.2529 (0.7817)	-0.4823 (0.7910)	-0.0369 (0.5192)
d1999	0.7314 (0.5147)	-0.0426 (0.7699)	0.0450 (0.6932)	0.1404 (0.6967)	-0.0588 (0.7147)	0.5121 (0.5352)
d2000	0.2791 (0.5274)	-0.5257 (0.7576)	-0.3430 (0.6909)	-0.2605 (0.6994)	-0.4311 (0.7035)	0.1170 (0.5053)
d2001	0.2860 (0.5467)	-0.4726 (0.7874)	-0.3902 (0.6864)	-0.2958 (0.6870)	-0.4996 (0.7106)	0.2906 (0.5413)
d2002	0.6547 (0.5160)	-0.1118 (0.7668)	-0.0725 (0.6690)	0.0313 (0.6707)	-0.1796 (0.6957)	0.3992 (0.5230)
d2003	-0.7429 (0.5528)	-1.4991 (0.8089)	-1.4764 (0.7229)	-1.3829 (0.7219)	-1.5897 (0.7495)	-0.8103 (0.5639)
d2004	-1.3795 (0.8338)	-2.0435 (0.9771)	-2.0304 (0.8766)	-1.9079 (0.8666)	-2.1790 (0.9084)	-1.7845 (0.8240)
d2005	-0.6369 (0.7284)	-1.5874 (0.8865)	-1.4940 (0.7936)	-1.3813 (0.8012)	-1.6298 (0.8078)	-1.3113 (0.7302)
d01Laval	0.8981 (0.3828)	0.8081 (0.3569)	0.5071 (0.3319)	0.5740 (0.3310)	0.4220 (0.3358)	1.9642 (0.4135)
d04MtIG	1.6467 (0.3681)	1.5229 (0.3647)	1.3400 (0.3270)	1.3825 (0.3293)	1.2356 (0.3128)	2.1442 (0.3720)
d07Concordia	-0.0798 (0.5227)	-0.0005 (0.5474)	-0.0851 (0.4894)	-0.0932 (0.4806)	-0.0876 (0.5013)	0.0469 (0.5755)
d09UQAMG	1.1200 (0.3701)	1.0910 (0.3805)	0.8237 (0.3232)	0.8340 (0.3212)	0.7631 (0.3225)	2.3426 (0.4285)
d08SherbrookeG	0.9489 (0.4823)	1.0250 (0.5345)	0.6464 (0.4833)	0.6685 (0.4778)	0.5820 (0.4881)	1.6192 (0.5540)
L_nbart_pre		0.1991 * (0.0862)				
L_subv_pre			0.1043 (0.1521)			
L_subv1_pre				0.0514 (0.1334)		
L_subv2_pre					0.1944 (0.1636)	
L_cont_pre						-0.3367 * (0.0843)
_cons	-6.3285 (1.4753)	-6.0768 (1.0252)	-7.2619 (2.1059)	-6.8564 (1.9593)	-7.9916 (2.1746)	-4.0137 (1.1105)

zinb du bloc C – breveter avec citations						
Variable	zinb_vce	zinb_art	zinb_subv	zinb_subv1	zinb_subv2	zinb_cont
Endogénéité	x	x	x	x	x	x
inflate						
L_nbarticlemoyan3	-0.3987 (0.2918)					
L_ltotsubvmoyan3	-1.5896 (0.7985)					-0.7375 (0.3930)
L_sq_ltotsubvmoyan3	0.1330 * (0.0698)					0.0439 ** (0.0317)
L_ltotcontmoyan3	-0.2199 (0.1602)	-0.0596 * (0.0622)	-0.0665 * (0.0697)	-0.0602 * (0.0593)	-0.0680 * (0.0705)	
L_lmcliqind23	0.1672 (0.8890)	-0.3117 (0.7026)	-0.1360 (0.8260)	-0.0830 (0.7506)	-0.1942 (0.8474)	-0.8855 (0.5956)
L_sq_lmcliqind23	-0.0223 ** (0.0273)	0.0016 ** (0.0233)	-0.0037 ** (0.0299)	-0.0056 ** (0.0276)	-0.0018 ** (0.0301)	0.0324 ** (0.0232)
deg_app0	0.2432 (0.2882)	-0.1206 (0.1899)	-0.1231 (0.2217)	-0.1322 (0.1816)	-0.1206 (0.2224)	-0.1096 (0.2114)
L_o_patpast	-7.0018 (5.9518)					
nano_excl	-0.8934 (1.8587)	-13.1823 (8.1054)	-2.9525 (18.2580)	-12.7853 (3.2136)	-2.5527 (12.0541)	-0.7841 (1.8403)
bio_excl	1.4270 (1.0315)	0.3924 (0.7454)	0.3539 (0.7353)	0.3347 (0.7091)	0.3429 (0.7481)	0.5336 (0.7736)
avg_pat_past		-37.8424 (28.2490)	-35.9839 (22.9351)	-34.4951 (19.3663)	-35.5689 (22.3720)	-126.0895 (310.7909)
_cons	0.7775 (2.8172)	1.3693 (1.8332)	1.1147 (2.0812)	1.0532 (1.9103)	1.2266 (2.1242)	4.7561 (1.7267)
Inalpha						
_cons	1.9175 (0.1153)	1.9006 (0.1243)	1.8837 (0.1290)	1.8606 (0.1259)	1.8851 (0.1296)	1.7973 (0.1214)
Statistics						
N	8946	8771	8775	8935	8775	8772
df_m	25	23	23	23	23	24
k	38	33	33	33	33	35
k_aux	1	1	1	1	1	1
k_eq	3	3	3	3	3	3
k_eq_model	1	1	1	1	1	1
k_dv	1	1	1	1	1	1
rank	38	33	33	33	33	35
ll	-1209.64	-1195.31	-1200.52	-1216.64	-1199.33	-1180.057
ll_0	-1591.17	-1539.25	-1539.29	-1560.47	-1539.29	-1543.897
chi2	564.3082	530.9975	464.3629	523.6008	470.6993	416.84554
aic	2495.278	2456.619	2467.034	2499.278	2464.657	2430.1145
bic	2765.038	2690.233	2700.663	2733.503	2698.286	2677.8907
p	3.4E-103	1.23E-97	8.92E-84	4.29E-96	4.33E-85	2.593E-73
ic	11	14	13	12	15	12
rc	0	0	0	0	0	0

Tableau P-4: Régressions tobit (deuxième étape) du bloc C – breveter avec citations

tobit du bloc C – breveter avec citations									
Variable	HSM_tobit_art		HSM_tobit_subv		HSM_tobit_subv_1		HSM_tobit_subv_2		HSM_tobit_cont
Endogénéité	x		x		x		x		x
model									
L_itotcontmoyan3	0.0594	***	0.0401		0.0466	*	0.0455	*	0.0354
	(0.0206)		(0.0253)		(0.0254)		(0.0249)		(0.1070)
L_lmliquind23	0.2576		0.4645	**	0.4496	**	0.4575	**	0.4188
	(0.2141)		(0.1941)		(0.1929)		(0.1926)		(0.1911)
L_sq_lmliquind23	-0.0102		-0.0188	**	-0.0182	**	-0.0186	**	-0.0172
	(0.0087)		(0.0078)		(0.0077)		(0.0077)		(0.0077)
deg_app0	0.0531		0.0851		0.0854		0.0897		0.1075
	(0.0651)		(0.0622)		(0.0611)		(0.0622)		(0.0605)
L_avg_pat_past	1.0722	***	1.1321	***	1.1516	***	1.1376	***	1.1865
	(0.3213)		(0.3234)		(0.3288)		(0.3242)		(0.3274)
nano_excl	-0.6741		-0.7284		-0.4068		-0.7405		-0.7964
	(0.5907)		(0.5904)		(0.5588)		(0.5875)		(0.5863)
bio_excl	-0.9243	***	-0.9943	***	-1.0308	***	-1.0064	***	-1.0477
	(0.2663)		(0.2584)		(0.2583)		(0.2584)		(0.2553)
L_nbarticlemoyan3	0.1390	*							
	(0.0716)								
L_nbart_res	-0.0336								
	(0.0872)								
d1997	-0.6385	*	-0.0141		-0.0274		-0.1249		-0.5907
	(0.3517)		(0.4648)		(0.4580)		(0.4632)		(0.3580)
d1998	-0.4033		-0.3322		-0.2412		-0.3342		-0.3274
	(0.3400)		(0.3395)		(0.3340)		(0.3396)		(0.3368)
d1999	-0.3606		-0.3145		-0.2560		-0.2992		-0.2230
	(0.3322)		(0.3265)		(0.3241)		(0.3262)		(0.3263)
d2000	-0.3120		-0.2072		-0.1481		-0.1973		-0.1303
	(0.3519)		(0.3431)		(0.3393)		(0.3424)		(0.3461)
d2001	-0.8491	**	-0.7561	**	-0.7297	*	-0.7442	*	-0.6666
	(0.3971)		(0.3840)		(0.3836)		(0.3832)		(0.3819)
d2002	-0.8105	**	-0.7512	*	-0.7096	*	-0.7301	*	-0.6247
	(0.3962)		(0.3880)		(0.3871)		(0.3874)		(0.3786)
d2003	-1.6527	***	-1.6332	***	-1.5947	***	-1.6136	***	-1.5058
	(0.4281)		(0.4280)		(0.4288)		(0.4273)		(0.4172)
d2004	-3.0308	***	-3.0030	***	-2.9597	***	-2.9796	***	-2.8523
	(0.5470)		(0.5415)		(0.5419)		(0.5408)		(0.5404)
d2005	-2.7968	***	-2.7534	***	-2.7018	***	-2.7253	***	-2.5921
	(0.5530)		(0.5592)		(0.5585)		(0.5594)		(0.5499)
d01Laval	0.2684		0.0444		0.0410		0.0674		0.2475
	(0.3587)		(0.3471)		(0.3468)		(0.3469)		(0.4099)
d04MtIG	0.3488		0.1154		0.1360		0.1323		0.2728
	(0.3206)		(0.3013)		(0.3011)		(0.3007)		(0.3611)
d07Concordia	-0.5400		-0.6114		-0.6367		-0.6271		-0.6835
	(0.8575)		(0.8576)		(0.8533)		(0.8569)		(0.8485)
d09UQAMG	0.2644		0.0945		0.0583		0.0986		0.1980
	(0.3876)		(0.3883)		(0.3863)		(0.3890)		(0.5018)
d08SherbrookeG	0.2540		-0.0332		-0.0782		-0.0091		0.1581
	(0.5411)		(0.5468)		(0.5490)		(0.5450)		(0.5575)
L_itotsubvmoyan3			0.2695		0.2663		0.2162		-0.0136
			(0.1713)		(0.1682)		(0.1680)		(0.1201)
L_sq_itotsubvmoyan3			-0.0060		-0.0064		-0.0048		0.0010
			(0.0081)		(0.0081)		(0.0081)		(0.0113)
L_subv_res			-0.2272	*					
			(0.1265)						
L_subv_res1					-0.2255	*			
					(0.1234)				
L_subv_res2							-0.1838		
							(0.1252)		



tobit du bloc C – breveter avec citations									
Variable	HSM_tobit_art		HSM_tobit_subv		HSM_tobit_subv_1		HSM_tobit_subv_2		HSM_tobit_cont
Endogénéité	x		x		x		x		x
L_cont_res									0.0361 (0.1059)
_cons	-5.9892	***	-8.1439	***	-8.0885	***	-7.7490	***	-6.0841 *** (0.7413)
	(0.6973)		(1.3587)		(1.3314)		(1.3378)		
sigma									
_cons	2.9155	***	2.9178	***	2.9116	***	2.9196	***	2.9201 *** (0.1086)
	(0.1131)		(0.1124)		(0.1120)		(0.1121)		
Statistics									
N	8771		8769		8929		8769		8769
df_m	23		24		24		24		24
k_aux	1		1		1		1		1
rank	25		26		26		26		26
ll	-1321.91		-1325.42		-1344.15		-1326.22		-1327.43
ll_0	-1470.23		-1470.17		-1494.12		-1470.17		-1470.17
aic	2693.821		2702.842		2740.304		2704.45		2706.868
bic	2870.802		2886.895		2924.827		2888.503		2890.921
r2_p	0.100883		0.098459		0.100372		0.097912		0.09709
p	1.16E-14		9.53E-14		1.05E-14		1.85E-13		7.53E-13

## ANNEXE Q – Régressions de deuxième étape du bloc D

Tableau Q-1: Régressions à binomiale négative (deuxième étape) du bloc D – nombre de brevets

nbreg du Bloc D– nombre de brevets									
Variable	nbreg_vce	nbreg_art	nbreg_subv	nbreg_subv1	nbreg_subv2	nbreg_cont			
Endogénéité	x		x	x	x	x	x		
L1.nbarticlemoyan3	-0.0021 (0.0223)	-0.0088 (0.0243)					-0.0110 (0.0311)		
L1.sq_nbarticlemo3	-0.0013 (0.0008)	-0.0016 (0.0009)	*				-0.0010 (0.0011)		
L1.ltotcontmoyan3	-0.0148 ** (0.0062)	-0.0138 ** (0.0060)	-0.0049 (0.0081)	-0.0045 (0.0084)	-0.0049 (0.0081)	-0.0551 *** (0.0175)			
L2.lmliquind23	0.2000 ** (0.0847)	0.2194 ** (0.0865)	0.1467 * (0.0777)	0.1456 * (0.0776)	0.1467 * (0.0777)	0.1498 (0.1065)			
L2.sq_lmliquind23	-0.0080 ** (0.0034)	-0.0088 ** (0.0035)		-0.0059 * (0.0031)	-0.0060 * (0.0031)	-0.0062 (0.0042)			
L1.deg_app0	0.0020 (0.0297)	0.0074 (0.0320)	0.0010 (0.0293)	0.0013 (0.0295)	0.0010 (0.0293)	0.0207 (0.0400)			
L1.avg_pat_past	0.2182 *** (0.0189)	0.2202 *** (0.0172)	0.2311 *** (0.0192)	0.2320 *** (0.0191)	0.2311 *** (0.0192)	0.2293 *** (0.0145)			
L1.o_patpast	0.1536 *** (0.0375)	0.1633 *** (0.0389)	0.1330 *** (0.0350)	0.1328 *** (0.0351)	0.1330 *** (0.0350)	0.1946 *** (0.0448)			
L1.nbloop0	0.0916 *** (0.0088)		0.1033 *** (0.0105)	0.1033 *** (0.0107)	0.1033 *** (0.0105)				
codechair33	-0.0188 (0.0328)	-0.0162 (0.0331)	0.0077 (0.0354)	0.0089 (0.0360)	0.0077 (0.0354)	-0.0183 (0.0348)			
age	-0.0876 (0.0748)	-0.0855 (0.0753)	-0.0957 (0.0758)	-0.0962 (0.0761)	-0.0957 (0.0758)	-0.1226 * (0.0705)			
sq_age	0.0025 (0.0026)	0.0024 (0.0027)	0.0031 (0.0027)	0.0031 (0.0028)	0.0031 (0.0027)	0.0034 (0.0024)			
d1997	-0.2687 ** (0.1334)	-0.2668 ** (0.1359)	-0.5118 *** (0.1980)	-0.5205 ** (0.2028)	-0.5118 *** (0.1980)	-0.3502 ** (0.1372)			
d1998	-0.0855 (0.1602)	-0.0927 (0.1595)	-0.0909 (0.1545)	-0.0923 (0.1544)	-0.0909 (0.1545)	-0.0643 (0.1586)			
d1999	-0.0040 (0.1686)	-0.0109 (0.1678)	-0.0345 (0.1671)	-0.0368 (0.1657)	-0.0345 (0.1671)	0.0011 (0.1785)			
d2000	-0.1654 (0.1238)	-0.1745 (0.1250)	-0.1267 (0.1184)	-0.1267 (0.1191)	-0.1267 (0.1184)	-0.1180 (0.1300)			
d2001	-0.1652 (0.1483)	-0.1695 (0.1499)	-0.2328 (0.1496)	-0.2358 (0.1499)	-0.2328 (0.1496)	-0.1221 (0.1532)			
d2002	0.1540 (0.2020)	0.1532 (0.2041)	0.1327 (0.1990)	0.1315 (0.1997)	0.1327 (0.1990)	0.1333 (0.1897)			
d2003	-0.1870 (0.1789)	-0.1874 (0.1812)	-0.2177 (0.1779)	-0.2192 (0.1784)	-0.2177 (0.1779)	-0.1855 (0.1701)			
d2004	-0.2213 (0.1692)	-0.2315 (0.1722)	-0.2474 (0.1676)	-0.2498 (0.1685)	-0.2474 (0.1676)	-0.2816 (0.1794)			
d2005	-0.0886 (0.2999)	-0.1025 (0.3055)	-0.1068 (0.3049)	-0.1092 (0.3067)	-0.1068 (0.3049)	-0.1751 (0.3077)			
d01Laval	-0.3893 *** (0.1419)	-0.3977 *** (0.1421)	-0.2758 ** (0.1302)	-0.2723 ** (0.1303)	-0.2758 ** (0.1302)	-0.4596 ** (0.1868)			
d04MtiG	-0.1802 (0.1400)	-0.1995 (0.1414)	-0.0920 (0.1301)	-0.0902 (0.1295)	-0.0920 (0.1301)	-0.2411 (0.1786)			
d07Concordia	0.0563 (0.1286)	0.0383 (0.1285)	0.0136 (0.1347)	0.0108 (0.1359)	0.0136 (0.1347)	-0.2189 (0.1983)			
d09UQAMG	-0.1400 (0.1533)	-0.1586 (0.1492)	-0.1067 (0.1527)	-0.1070 (0.1528)	-0.1067 (0.1527)	-0.0836 (0.1964)			
d08SherbrookeG	-0.2893 * (0.1510)	-0.3070 ** (0.1500)	-0.1827 (0.1456)	-0.1800 (0.1457)	-0.1827 (0.1456)	-0.3804 ** (0.1860)			
L_nbloop0		0.0894 *** (0.0087)							
L_nbart_res		0.0238 (0.0224)							
L2_sq_lmliquind23			-0.0060 * (0.0031)						
L1.ltotsubvmoyan3			-0.0781 * (0.0031)	-0.0809 * (0.0031)	-0.0781 * (0.0031)				

nbreg du Bloc D– nombre de brevets						
Variable	nbreg_vce	nbreg_art	nbreg_subv	nbreg_subv1	nbreg_subv2	nbreg_cont
Endogénéité	x	x	x	x	x	x
L2.ltotsubvmoyan3			(0.0409)	(0.0436)	(0.0409)	0.0227 *
						(0.0133)
L_subv_res			0.0813 *			
			(0.0435)			
L_subv_res1				0.0842 *		
				(0.0469)		
L_subv_res2					0.0813 *	
					(0.0435)	
L_cont_res						0.0448 **
						(0.0185)
_cons	0.7785	0.7263	1.6203 **	1.6495 **	1.6203 **	1.1929 ***
	(0.4786)	(0.4914)	(0.7229)	(0.7459)	(0.7229)	(0.4512)
Inalpha						
_cons	-3.53032 ***	-3.55867 ***	-3.47405 ***	-3.47993 ***	-3.47405 ***	-3.25657 ***
	1.256137	1.268192	1.185045	1.179618	1.185045	1.104259
Statistics						
N	526	526	526	526	526	526
df_m	26	27	26	26	26	27
k	28	29	28	28	28	29
k_aux	1	1	1	1	1	1
k_eq	2	2	2	2	2	2
k_eq_model	1	1	1	1	1	1
k_dv	1	1	1	1	1	1
ll	-733.389	-733.041	-733.931	-733.872	-733.931	-743.258
ll_0	-835.987	-835.987	-835.987	-835.987	-835.987	-835.987
chi2	3601.597	3820.041	3075.949	3085.047	3075.949	4312.141
aic	1522.777	1524.081	1523.862	1523.743	1523.862	1544.516
bic	1642.206	1647.775	1643.29	1643.172	1643.29	1668.21
p	0	0	0	0	0	0

Tableau Q-2: Régressions tobit (deuxième étape) du bloc D – nombre de brevets

tobit du Bloc D – nombre de brevets									
Variable	HSM_tobit_art	HSM_tobit_subv	HSM_tobit_subv_1	HSM_tobit_subv_2	HSM_tobit_cont				
Endogénéité	x	x	x	x	x				
model									
L_ltotcontmoyan3	-0.0110 (0.0136)	0.0090 (0.0185)	0.0099 (0.0187)	0.0090 (0.0185)	-0.0421 (0.0257)				
L2_lmcliqind23	0.2974 (0.1960)	0.2329 (0.1820)	0.2312 (0.1824)	0.2329 (0.1820)	0.2257 (0.1909)				
L2_sq_lmcliqind23	-0.0115 (0.0077)	-0.0091 (0.0072)	-0.0091 (0.0072)	-0.0091 (0.0072)	-0.0088 (0.0075)				
L_deg_app0	0.0115 (0.0604)	0.0254 (0.0556)	0.0261 (0.0556)	0.0254 (0.0556)	0.0019 (0.0573)				
L_avg_pat_past	0.3182 *** (0.0374)	0.3349 *** (0.0423)	0.3363 *** (0.0427)	0.3349 *** (0.0423)	0.3306 *** (0.0358)				
L_o_patpast	0.3842 *** (0.0968)	0.3594 *** (0.0919)	0.3591 *** (0.0920)	0.3594 *** (0.0919)	0.3553 *** (0.0941)				
L_nbloop0	0.1405 *** (0.0307)	0.1562 *** (0.0304)	0.1566 *** (0.0305)	0.1562 *** (0.0304)	0.1502 *** (0.0308)				
codechair33	-0.0502 (0.0771)	0.0116 (0.0854)	0.0138 (0.0857)	0.0116 (0.0854)	-0.0299 (0.0780)				
age	-0.0106 (0.1394)	-0.0340 (0.1389)	-0.0342 (0.1392)	-0.0340 (0.1389)	-0.0374 (0.1393)				
sq_age	-0.0004 (0.0053)	0.0011 (0.0054)	0.0012 (0.0054)	0.0011 (0.0054)	0.0007 (0.0053)				
L_nbarticlemoyan3	0.0488 (0.0596)				0.0728 (0.0564)				
L_sq_nbart~3	-0.0048 ** (0.0024)				-0.0047 ** (0.0023)				
L_nbart_res	0.0366 (0.0494)								
d1997	-0.6297 * (0.3280)	-1.0716 *** (0.3965)	-1.0877 *** (0.3999)	-1.0716 *** (0.3965)	-0.6745 ** (0.3314)				
d1998	-0.8080 ** (0.3283)	-0.7864 ** (0.3178)	-0.7892 ** (0.3177)	-0.7864 ** (0.3178)	-0.7726 ** (0.3301)				
d1999	-0.4992 (0.3171)	-0.5231 * (0.3164)	-0.5259 * (0.3165)	-0.5231 * (0.3164)	-0.4380 (0.3186)				
d2000	-0.4614 * (0.2670)	-0.3463 (0.2507)	-0.3451 (0.2506)	-0.3463 (0.2507)	-0.4008 (0.2658)				
d2001	-0.7452 ** (0.3599)	-0.8255 ** (0.3531)	-0.8303 ** (0.3535)	-0.8255 ** (0.3531)	-0.7084 ** (0.3584)				
d2002	-0.1585 (0.3610)	-0.1646 (0.3515)	-0.1675 (0.3519)	-0.1646 (0.3515)	-0.1212 (0.3602)				
d2003	-0.6698 * (0.3880)	-0.7118 * (0.3801)	-0.7143 * (0.3804)	-0.7118 * (0.3801)	-0.6344 (0.3858)				
d2004	-0.6749 (0.4208)	-0.7375 * (0.4045)	-0.7412 * (0.4052)	-0.7375 * (0.4045)	-0.6903 * (0.4156)				
d2005	-0.7131 (0.5404)	-0.7163 (0.5385)	-0.7184 (0.5397)	-0.7163 (0.5385)	-0.6949 (0.5432)				
d01Laval	-0.4511 * (0.2509)	-0.2600 (0.2593)	-0.2528 (0.2617)	-0.2600 (0.2593)	-0.3755 (0.2634)				
d04MtlG	-0.2517 (0.2362)	-0.1190 (0.2278)	-0.1150 (0.2292)	-0.1190 (0.2278)	-0.1787 (0.2359)				
d07Concordia	0.7573 ** (0.2960)	0.6663 ** (0.3094)	0.6606 ** (0.3098)	0.6663 ** (0.3094)	0.7008 ** (0.3210)				
d09UQAMG	-0.1366 (0.2993)	-0.0970 (0.3012)	-0.0963 (0.3013)	-0.0970 (0.3012)	0.0054 (0.3211)				
d08SherbrookeG	-0.4374 (0.3193)	-0.2951 (0.3191)	-0.2910 (0.3202)	-0.2951 (0.3191)	-0.3754 (0.3229)				
L_ltotsubvmoyan3		-0.1413 * (0.0819)	-0.1470 * (0.0837)	-0.1413 * (0.0819)					
L_subv_res		0.1313 (0.0875)							
L_subv_res1			0.1373 (0.0895)						
L_subv_res2				0.1313 (0.0875)					
L2_ltotsubvmoyan3					-0.0027 (0.0230)				

tobit du Bloc D – nombre de brevets									
Variable	HSM_tobit_art		HSM_tobit_subv		HSM_tobit_subv_1		HSM_tobit_subv_2		HSM_tobit_cont
Endogénéité	x		x		x		x		x
L_cont_res									0.0367 (0.0286)
_cons	-1.0116 (0.9961)		0.5124 (1.3372)		0.5670 (1.3482)		0.5124 (1.3372)		-0.6449 (0.9771)
sigma									
_cons	1.2080 (0.0722)	***	1.2089 (0.0707)	***	1.2088 (0.0708)	***	1.2089 (0.0707)	***	1.2039 (0.0723) ***
Statistics									
N	526		526		526		526		526
df_m	27		26		26		26		28
k_aux	1		1		1		1		1
rank	29		28		28		28		30
ll	-380.955		-381.053		-380.987		-381.053		-380.555
ll_0	-430.751		-430.751		-430.751		-430.751		-430.751
aic	819.9105		818.1056		817.9744		818.1056		821.1092
bic	943.6043		937.5341		937.4029		937.5341		949.0682
r2_p	0.115602		0.115375		0.115528		0.115375		0.116532
p	3.88E-40		1.89E-35		7.91E-36		1.89E-35		2.13E-50

## ANNEXE R – Régressions de deuxième étape du bloc E

Tableau R-1: Régressions à binomiale négative (deuxième étape) du bloc E – nombre de revendications

Nbreg du Bloc E – nombre de revendications												
Variable	nbreg_vce	nbreg_art	nbreg_vce1	nbreg_subv	nbreg_subv1	nbreg_subv2	nbreg_vce2	nbreg_cont				
Endogénéité	x		x		x		x		x			
<b>L1.nbarticlemoyan3</b>	0.0086 (0.0188)	0.0289 (0.0246)										
<b>L1.ltotcontmoyan3</b>	-0.0661 * (0.0352)	-0.0651 * (0.0348)	-0.0666 * (0.0342)	-0.0819 ** (0.0351)	-0.0779 ** (0.0348)	-0.0818 ** (0.0353)	-0.0666 * (0.0342)	0.0411 (0.0636)				
<b>L1.sq_ltotcontmoyan3</b>	0.0068 ** (0.0027)	0.0066 ** (0.0027)	0.0071 *** (0.0026)	0.0070 *** (0.0026)	0.0068 *** (0.0026)	0.0066 *** (0.0025)	0.0071 *** (0.0026)	0.0038 (0.0033)				
<b>L1.lmcliquind23</b>	0.2040 (0.1511)	0.1724 (0.1617)	0.2203 (0.1414)	0.2406 * (0.1405)	0.2298 * (0.1394)	0.2425 * (0.1393)	0.2203 (0.1414)	0.2211 (0.1392)				
<b>L1.sq_lmcliquind23</b>	-0.0080 (0.0060)	-0.0066 (0.0064)	-0.0086 (0.0056)	-0.0096 * (0.0055)	-0.0092 * (0.0055)	-0.0097 * (0.0055)	-0.0086 (0.0055)	-0.0083 (0.0055)				
<b>deg_app</b>	0.0228 (0.0791)	0.0171 (0.0782)	0.0312 (0.0811)	-0.0060 (0.0836)	-0.0022 (0.0836)	-0.0133 (0.0831)	0.0312 (0.0811)	0.0380 (0.0815)				
<b>nbloop</b>	0.1688 *** (0.0314)	0.1737 *** (0.0319)	0.1717 *** (0.0303)	0.1551 *** (0.0329)	0.1556 *** (0.0322)	0.1503 *** (0.0333)	0.1717 *** (0.0303)	0.1762 *** (0.0298)				
<b>propuninv</b>	-0.7860 *** (0.2719)	-0.8134 *** (0.2693)	-0.7522 *** (0.2766)	-0.8854 *** (0.2817)	-0.8489 *** (0.2831)	-0.9042 *** (0.2845)	-0.7522 *** (0.2766)	-0.7715 *** (0.2767)				
<b>excl_nano_pat</b>	0.6254 *** (0.1495)	0.6117 *** (0.1496)	0.6181 *** (0.1478)	0.6435 *** (0.1507)	0.6510 *** (0.1509)	0.6566 *** (0.1499)	0.6181 *** (0.1478)	0.6680 *** (0.1532)				
<b>excl_bio_pat</b>	0.1626 (0.1171)	0.1990 * (0.1102)	0.1281 (0.1138)	0.2145 * (0.1254)	0.1958 (0.1249)	0.2372 * (0.1241)	0.1281 (0.1138)	0.0947 (0.1108)				
<b>codechair33</b>	-0.0460 (0.0609)	-0.0495 (0.0596)	-0.0266 (0.0642)	-0.0655 (0.0652)	-0.0609 (0.0653)	-0.0696 (0.0654)	-0.0266 (0.0642)	-0.0374 (0.0629)				
<b>age</b>	0.2117 *** (0.0716)	0.2069 ** (0.0813)	0.2137 *** (0.0696)	0.2049 *** (0.0789)	0.2008 *** (0.0709)	0.1984 ** (0.0777)	0.2137 *** (0.0696)	0.2686 *** (0.0757)				
<b>sq_age</b>	-0.0076 *** (0.0027)	-0.0076 ** (0.0031)	-0.0075 *** (0.0027)	-0.0077 ** (0.0030)	-0.0074 *** (0.0028)	-0.0076 ** (0.0030)	-0.0075 *** (0.0027)	-0.0097 *** (0.0029)				
<b>d1997</b>	-0.1870 (0.2596)	-0.1821 (0.2590)	-0.2600 (0.2531)	0.0037 (0.2607)	-0.0329 (0.2653)	0.0683 (0.2530)	-0.2600 (0.2531)	-0.3382 (0.2512)				
<b>d1998</b>	-0.1987 (0.2563)	-0.1767 (0.2559)	-0.1954 (0.2554)	-0.1233 (0.2504)	-0.1464 (0.2517)	-0.1147 (0.2461)	-0.1954 (0.2554)	-0.3335 (0.2553)				
<b>d1999</b>	-0.2663 (0.2534)	-0.2757 (0.2525)	-0.2562 (0.2512)	-0.1403 (0.2513)	-0.1761 (0.2498)	-0.1131 (0.2500)	-0.2562 (0.2512)	-0.4451 * (0.2615)				
<b>d2000</b>	-0.0404 (0.2629)	-0.0483 (0.2635)	0.0009 (0.2641)	-0.0384 (0.2550)	-0.0491 (0.2574)	-0.0317 (0.2508)	0.0009 (0.2641)	-0.1464 (0.2698)				
<b>d2001</b>	-0.2311 (0.2833)	-0.2371 (0.2815)	-0.2353 (0.2799)	-0.1164 (0.2781)	-0.1478 (0.2797)	-0.0909 (0.2740)	-0.2353 (0.2799)	-0.4367 (0.2926)				

Nbreg du Bloc E – nombre de revendications								
Variable	nbreg_vce	nbreg_art	nbreg_vce1	nbreg_subv	nbreg_subv1	nbreg_subv2	nbreg_vce2	nbreg_cont
<b>Endogénéité</b>		<b>x</b>		<b>x</b>	<b>x</b>	<b>x</b>		<b>x</b>
<b>d2002</b>	-0.1046 (0.3016)	-0.0757 (0.2990)	-0.0948 (0.3019)	-0.0581 (0.2938)	-0.0863 (0.2959)	-0.0365 (0.2911)	-0.0948 (0.3019)	-0.2930 (0.3173)
<b>d2003</b>	-0.3116 (0.2759)	-0.2915 (0.2719)	-0.3104 (0.2748)	-0.2353 (0.2696)	-0.2644 (0.2727)	-0.2134 (0.2664)	-0.3104 (0.2748)	-0.5004 (0.2926) *
<b>d2004</b>	-0.3639 (0.2900)	-0.3292 (0.2878)	-0.3911 (0.2783)	-0.3143 (0.2672)	-0.3472 (0.2701)	-0.2885 (0.2633)	-0.3911 (0.2783)	-0.4900 (0.2850) *
<b>d2005</b>	-0.2577 (0.3505)	-0.2572 (0.3490)	-0.2485 (0.3493)	-0.2601 (0.3368)	-0.2788 (0.3403)	-0.2574 (0.3327)	-0.2485 (0.3493)	-0.3657 (0.3526)
<b>d01Laval</b>	-0.0666 (0.1660)	-0.0301 (0.1643)	-0.0609 (0.1644)	-0.1645 (0.1642)	-0.1616 (0.1664)	-0.2076 (0.1670)	-0.0609 (0.1644)	-0.0879 (0.1631)
<b>d04MtlG</b>	0.0705 (0.1766)	0.0981 (0.1723)	0.0755 (0.1794)	-0.0229 (0.1774)	-0.0103 (0.1806)	-0.0604 (0.1768)	0.0755 (0.1794)	0.0359 (0.1772)
<b>d07Concordia</b>	0.3215 (0.3243)	0.3816 (0.3249)	0.3217 (0.3225)	0.3943 (0.3202)	0.3908 (0.3206)	0.3943 (0.3206)	0.3217 (0.3225)	0.5118 (0.3317)
<b>d09UQAMG</b>	0.0625 (0.2551)	0.0967 (0.2505)	0.0524 (0.2572)	0.0292 (0.2423)	0.0299 (0.2444)	0.0127 (0.2392)	0.0524 (0.2572)	-0.1457 (0.2428)
<b>d08SherbrookeG</b>	0.1462 (0.2389)	0.1953 (0.2336)	0.1206 (0.2388)	0.0716 (0.2423)	0.0699 (0.2407)	0.0438 (0.2449)	0.1206 (0.2388)	0.1600 (0.2417)
<b>L_nbart_res</b>		-0.0401 (0.0324)						
<b>L1.ltotsubvmoyan3</b>			-0.0259 (0.0195)	0.0793 (0.0493)	0.0675 (0.0480)	0.1035 * (0.0532)	-0.0259 (0.0195)	-0.0653 ** (0.0268)
<b>L_subv_res</b>				-0.1189 ** 0.0532				
<b>L_subv_res1</b>					-0.1055 ** (0.0523)			
<b>L_subv_res2</b>						-0.1448 ** (0.0569)		
<b>L_cont_res</b>								-0.0731 ** (0.0314)
<b>_cons</b>	1.5953 ** (0.6440)	1.6355 ** (0.7192)	1.7623 *** (0.6370)	0.9754 (0.7697)	1.1196 (0.7070)	0.8231 (0.7814)	1.7623 *** (0.6370)	1.6690 *** (0.6411)
<b>Inalpha</b>								
<b>_cons</b>	-0.2990 (0.0715)	-0.3022 (0.0712)	-0.3033 (0.0712)	-0.3123 (0.0721)	-0.3102 (0.0725)	-0.3157 (0.0719)	-0.3033 (0.0712)	-0.3111 (0.0721)
<b>N</b>	422	419	422	419	422	419	422	422
<b>df_m</b>	27	28	27	28	28	28	27	28
<b>k</b>	29	30	29	30	30	30	29	30
<b>k_aux</b>	1	1	1	1	1	1	1	1
<b>k_eq</b>	2	2	2	2	2	2	2	2
<b>k_eq_model</b>	1	1	1	1	1	1	1	1
<b>k_dv</b>	1	1	1	1	1	1	1	1

Nbreg du Bloc E – nombre de revendications								
Variable	nbreg_vce	nbreg_art	nbreg_vce1	nbreg_subv	nbreg_subv1	nbreg_subv2	nbreg_vce2	nbreg_cont
Endogénéité	x		x		x		x	
ll	-1803.4729	-1790.4857	-1802.4751	-1788.3472	-1800.8311	-1787.3848	-1802.4751	-1800.6505
ll_0	-1859.1698	-1847.3972	-1859.1698	-1847.3972	-1859.1698	-1847.3972	-1859.1698	-1859.1698
chi2	155.40807	170.92575	171.0114	158.04505	155.01177	156.86404	171.0114	162.18531
aic	3664.9457	3640.9715	3662.9502	3636.6943	3661.6622	3634.7696	3662.9502	3661.301
bic	3782.2509	3762.1076	3780.2554	3757.8305	3783.0124	3755.9057	3780.2554	3782.6511
p	5.32E-20	1.88E-22	7.08E-23	4.31E-20	1.53E-19	7.06E-20	7.08E-23	7.57E-21



Tableau R-2: Régressions à Moindres Carrés Ordinaires (deuxième étape) du bloc E – nombre de revendications

Reg du Bloc E – nombre de revendications																	
Variable	reg_vce		reg_art		reg_vce1		reg_subv		reg_subv1		reg_subv2		reg_vce2		reg_cont		
Endogénéité			x				x		x		x				x		
L1.nbarticlemoyan3	0.0167		0.0327														
	(0.0170)		(0.0266)														
L1.ltotcontmoyan3	-0.0689	*	-0.0658	*	-0.0725	*	-0.0796	**	-0.0769	**	-0.0810	**	-0.0725	*		-0.0030	
	(0.0380)		(0.0389)		(0.0383)		(0.0386)		(0.0384)		(0.0390)		(0.0383)			(0.0804)	
L1.sq_ltotcontmoyan3	0.0073	**	0.0071	**	0.0079	***	0.0077	**	0.0075	**	0.0076	**	0.0079	***		0.0061	
	(0.0030)		(0.0031)		(0.0030)		(0.0030)		(0.0030)		(0.0030)		(0.0030)			(0.0038)	
L1.lmcliquind23	0.0656		0.0353		0.1013		0.1169		0.1082		0.1191		0.1013			0.1066	
	(0.1591)		(0.1735)		(0.1459)		(0.1477)		(0.1461)		(0.1482)		(0.1459)			(0.1449)	
L1.sq_lmcliquind23	-0.0020		-0.0007		-0.0034		-0.0042		-0.0038		-0.0043		-0.0034			-0.0034	
	(0.0063)		(0.0069)		(0.0057)		(0.0058)		(0.0057)		(0.0058)		(0.0057)			(0.0057)	
deg_app	-0.0833		-0.0856		-0.0801		-0.1067		-0.1001		-0.1122		-0.0801			-0.0701	
	(0.0817)		(0.0832)		(0.0826)		(0.0830)		(0.0821)		(0.0836)		(0.0826)			(0.0836)	
nbloop	0.1999	***	0.2031	***	0.2010	***	0.1894	***	0.1896	***	0.1858	***	0.2010	***		0.2059	***
	(0.0303)		(0.0305)		(0.0303)		(0.0314)		(0.0316)		(0.0316)		(0.0303)			(0.0324)	
propuninv	-0.4755	*	-0.5032	*	-0.4575	*	-0.5580	**	-0.5192	*	-0.5766	**	-0.4575	*		-0.4663	*
	(0.2559)		(0.2576)		(0.2594)		(0.2681)		(0.2645)		(0.2784)		(0.2594)			(0.2605)	
excl_nano_pat	0.6448	***	0.6350	***	0.6485	***	0.6635	***	0.6651	***	0.6682	***	0.6485	***		0.6774	***
	(0.1495)		(0.1492)		(0.1462)		(0.1497)		(0.1496)		(0.1493)		(0.1462)			(0.1559)	
excl_bio_pat	0.0852		0.1045		0.0592		0.1144		0.1014		0.1274		0.0592			0.0301	
	(0.1294)		(0.1281)		(0.1247)		(0.1399)		(0.1388)		(0.1406)		(0.1247)			(0.1197)	
codechair33	-0.0253		-0.0318		-0.0143		-0.0387		-0.0342		-0.0450		-0.0143			-0.0181	
	(0.0688)		(0.0694)		(0.0727)		(0.0739)		(0.0733)		(0.0772)		(0.0727)			(0.0722)	
age	0.0754		0.0681		0.0885		0.0891		0.0852		0.0866		0.0885			0.1235	
	(0.0763)		(0.0877)		(0.0719)		(0.0817)		(0.0726)		(0.0808)		(0.0719)			(0.0820)	
sq_age	-0.0029		-0.0027		-0.0032		-0.0035		-0.0033		-0.0035		-0.0032			-0.0046	
	(0.0030)		(0.0033)		(0.0029)		(0.0032)		(0.0029)		(0.0032)		(0.0029)			(0.0033)	
d1997	0.0164		0.0062		-0.0140		0.1366		0.1213		0.1782		-0.0140			-0.0512	
	(0.2684)		(0.2712)		(0.2645)		(0.2719)		(0.2757)		(0.2684)		(0.2645)			(0.2597)	
d1998	0.0327		0.0624		0.0302		0.0692		0.0382		0.0714		0.0302			-0.0517	
	(0.2480)		(0.2529)		(0.2469)		(0.2480)		(0.2443)		(0.2466)		(0.2469)			(0.2480)	
d1999	-0.0328		-0.0446		-0.0264		0.0146		0.0028		0.0257		-0.0264			-0.1448	
	(0.2428)		(0.2508)		(0.2433)		(0.2458)		(0.2418)		(0.2452)		(0.2433)			(0.2642)	
d2000	0.1245		0.1071		0.1547		0.1172		0.1179		0.1069		0.1547			0.0742	
	(0.2582)		(0.2641)		(0.2579)		(0.2592)		(0.2582)		(0.2599)		(0.2579)			(0.2647)	
d2001	-0.1442		-0.1541		-0.1439		-0.0895		-0.1008		-0.0750		-0.1439			-0.2582	
	(0.3002)		(0.3049)		(0.2987)		(0.3018)		(0.3006)		(0.2993)		(0.2987)			(0.3111)	
d2002	-0.0495		-0.0520		-0.0426		-0.0472		-0.0514		-0.0486		-0.0426			-0.1573	
	(0.2996)		(0.3036)		(0.3004)		(0.3007)		(0.2993)		(0.3010)		(0.3004)			(0.3114)	
d2003	-0.0478		-0.0411		-0.0562		-0.0325		-0.0409		-0.0269		-0.0562			-0.1439	

Reg du Bloc E – nombre de revendications													
Variable	reg_vce		reg_art		reg_vce1		reg_subv		reg_subv1		reg_subv2		reg_vce2
Endogénéité	x		x		x		x		x		x		x
	(0.2783)		(0.2787)		(0.2762)		(0.2749)		(0.2746)		(0.2734)		(0.2762)
d2004	-0.0694		-0.0581		-0.0800		-0.0474		-0.0604		-0.0408		-0.0800
	(0.2968)		(0.2977)		(0.2915)		(0.2902)		(0.2902)		(0.2888)		(0.2915)
d2005	0.0202		0.0115		0.0362		0.0123		0.0120		0.0058		0.0362
	(0.3961)		(0.3970)		(0.3936)		(0.3949)		(0.3944)		(0.3957)		(0.3936)
d01Laval	0.0281		0.0594		0.0295		-0.0375		-0.0427		-0.0600		0.0295
	(0.1714)		(0.1711)		(0.1786)		(0.1869)		(0.1898)		(0.1878)		(0.1786)
d04MtlG	0.1219		0.1464		0.1145		0.0519		0.0564		0.0348		0.1145
	(0.1869)		(0.1855)		(0.1921)		(0.1979)		(0.1986)		(0.1964)		(0.1921)
d07Concordia	0.3094		0.3508		0.2565		0.3063		0.2987		0.3191		0.2565
	(0.2730)		(0.2747)		(0.2745)		(0.2759)		(0.2764)		(0.2801)		(0.2745)
d09UQAMG	0.1492		0.1696		0.1309		0.1125		0.1145		0.1081		0.1309
	(0.2847)		(0.2855)		(0.2890)		(0.2839)		(0.2837)		(0.2828)		(0.2890)
d08SherbrookeG	0.1086		0.1463		0.0876		0.0453		0.0440		0.0317		0.0876
	(0.2680)		(0.2671)		(0.2727)		(0.2753)		(0.2757)		(0.2746)		(0.2727)
L_nbart_res			-0.0328										
			(0.0338)										
L1.ltotsubvmoyan3					-0.0175		0.0495		0.0422		0.0678		-0.0175
					(0.0230)		(0.0568)		(0.0564)		(0.0656)		(0.0230)
L_subv_res							-0.0752						
							0.0592						
L_subv_res1									-0.0659				
									(0.0583)				
L_subv_res2											-0.0943		
											(0.0676)		
L_cont_res													-0.0508
													(0.0445)
_cons	2.3073 ***		2.3842 ***		2.3296 ***		1.8223 **		1.9027 **		1.6983 *		2.3296 ***
	(0.7008)		(0.7756)		(0.6977)		(0.8463)		(0.7952)		(0.9033)		(0.6977)
Statistics													
N	422		419		422		419		422		419		422
df_m	27		28		27		28		28		28		27
k_aux													
rank	28		29		28		29		29		29		28
ll	-588.44982		-584.43168		-588.50359		-584.20272		-587.94992		-583.82688		-588.50359
ll_0	-621.5599		-618.32135		-621.5599		-618.32135		-621.5599		-618.32135		-621.5599
aic	1232.8996		1226.8634		1233.0072		1226.4054		1233.8998		1225.6538		1233.0072
bic	1346.1598		1343.9616		1346.2673		1343.5037		1351.205		1342.752		1346.2673
r2	0.14522739		0.1493585		0.14500953		0.15028769		0.14725011		0.15181068		0.14500953

Tableau R-3: Régressions à binomiale négative (deuxième étape) du bloc E – nombre de revendications avec sous-catégories de brevets

nbreg de Bloc E – nombre de revendications avec sous-catégories de brevets																
Variable	nbreg_vce		nbreg_art		nbreg_vce1		nbreg_subv		nbreg_subv1		nbreg_subv2		nbreg_vce2		nbreg_cont	
Endogénéité			x				x		x		x				x	
L1.nbarticlemoyan3	0,038391	**	0,062908	***												
	0,018295		0,024019													
L1.ltotcontmoyan3	-0,06801	**	-0,0653	**	-0,07577	**	-0,08547	***	-0,08343	**	-0,08387	**	-0,07577	**	0,079378	
	0,033532		0,033103		0,032978		0,032658		0,032728		0,032737		0,032978		0,06416	
L1.sq_ltotcontmoyan3	0,00657	**	0,006197	**	0,007524	***	0,007254	***	0,007093	***	0,007143	***	0,007524	***	0,002997	
	0,002714		0,002735		0,002619		0,002587		0,002609		0,002578		0,002619		0,003369	
L1.lmcliquind23	0,047944		0,000398		0,14152		0,147061		0,140208		0,149239		0,14152		0,150275	
	0,112558		0,123479		0,107716		0,106742		0,1054		0,107367		0,107716		0,103338	
L1.sq_lmcliquind23	-0,00147		0,000503		-0,00528		-0,00569		-0,00541		-0,00577		-0,00528		-0,00507	
	0,00445		0,004893		0,004222		0,004191		0,004139		0,004214		0,004222		0,004047	
deg_app	0,050323		0,045944		0,057973		0,032951		0,032473		0,033959		0,057973		0,053626	
	0,074825		0,074046		0,075903		0,077481		0,07693		0,077364		0,075903		0,075209	
nbloop	0,147909	***	0,151408	***	0,145366	***	0,126163	***	0,125865	***	0,126334	***	0,145366	***	0,156568	***
	0,02478		0,025839		0,025136		0,027915		0,027554		0,027704		0,025136		0,024434	
propuninv	-0,46186	**	-0,48799	**	-0,48134	**	-0,5845	***	-0,55557	***	-0,58349	***	-0,48134	**	-0,50188	**
	0,199825		0,197982		0,201415		0,20742		0,209037		0,208884		0,201415		0,198201	
metrology_processus_pat	-0,1808		-0,18447		-0,11449		-0,12629		-0,12193		-0,12437		-0,11449		0,05825	
	0,225491		0,222498		0,219503		0,219872		0,220563		0,219285		0,219503		0,217347	
devices_electronic_pat	0,474692	**	0,46022	**	0,457299	**	0,40273	**	0,409133	**	0,405348	**	0,457299	**	0,646683	***
	0,201846		0,200852		0,198783		0,205132		0,205136		0,204811		0,198783		0,212258	
Structures_materials_pat	0,620963	***	0,611606	***	0,594249	***	0,677932	***	0,68373	***	0,672323	***	0,594249	***	0,295656	
	0,179195		0,178985		0,171837		0,173332		0,17443		0,171961		0,171837		0,191988	
dna_rna_pat	0,18973		0,197315		0,186867		0,174345		0,181096		0,175159		0,186867		0,191823	
	0,127988		0,127195		0,132084		0,131562		0,131467		0,132249		0,132084		0,129138	
proteins_molecules_pat	0,348056	**	0,381745	**	0,331596	**	0,402452	**	0,399951	**	0,403145	**	0,331596	**	0,34163	**
	0,152559		0,15057		0,156143		0,159857		0,159163		0,158555		0,156143		0,152691	
cell_tissu_engineerin_pat	0,923019	***	0,957166	***	0,839883	***	0,87342	***	0,877211	***	0,867188	***	0,839883	***	0,686555	***
	0,154369		0,151971		0,153879		0,147969		0,148632		0,148799		0,153879		0,161371	
process_biotech_pat	0,470848	***	0,505737	***	0,436812	***	0,510939	***	0,492994	***	0,509567	***	0,436812	***	0,216401	
	0,127404		0,122116		0,126143		0,126678		0,127861		0,125437		0,126143		0,138523	
gene_rna_vector_pat	0,361238	***	0,350929	***	0,336624	**	0,38933	***	0,395095	***	0,384518	***	0,336624	**	0,411765	***
	0,128568		0,125491		0,135204		0,134595		0,135301		0,136067		0,135204		0,135759	

nbreg de Bloc E – nombre de revendications avec sous-catégories de brevets																
Variable	nbreg_vce		nbreg_art		nbreg_vce1		nbreg_subv		nbreg_subv1		nbreg_subv2		nbreg_vce2		nbreg_cont	
Endogénéité			x				x				x				x	
Bioinformatics_pat	0,450349	**	0,415378	*	0,51441	**	0,51371	**	0,520925	**	0,515531	**	0,51441	**	0,734188	***
	0,223661		0,217611		0,226335		0,21891		0,219074		0,219186		0,226335		0,240113	
codechair33	-0,06995		-0,07485		-0,04256		-0,07239		-0,07028		-0,06879		-0,04256		-0,0522	
	0,049251		0,049371		0,055224		0,05476		0,054641		0,056084		0,055224		0,053631	
age	0,109225	*	0,082824		0,14119	**	0,128994	*	0,132134	**	0,129477	*	0,14119	**	0,210033	***
	0,062852		0,072689		0,06128		0,07148		0,062323		0,071515		0,06128		0,064607	
sq_age	-0,00463	*	-0,00387		-0,00542	**	-0,00536	*	-0,00544	**	-0,00537	*	-0,00542	**	-0,00828	***
	0,002472		0,002779		0,002417		0,00276		0,00246		0,00275		0,002417		0,002513	
d1997	0,055833		0,046244		0,062678		0,256027		0,254935		0,247764		0,062678		-0,0826	
	0,195605		0,197249		0,200718		0,222923		0,224225		0,21844		0,200718		0,198661	
d1998	0,085267		0,116224		0,053004		0,116645		0,091187		0,107913		0,053004		-0,19236	
	0,194634		0,197538		0,196501		0,193758		0,191171		0,194169		0,196501		0,204285	
d1999	0,08815		0,062958		0,097445		0,17238		0,155041		0,164678		0,097445		-0,18247	
	0,222282		0,218413		0,225893		0,223457		0,22099		0,226113		0,225893		0,240841	
d2000	0,196435		0,17539		0,245631		0,20923		0,19574		0,214778		0,245631		0,00751	
	0,215306		0,217255		0,213867		0,208279		0,207499		0,20926		0,213867		0,221831	
d2001	0,030559		0,015811		0,0448		0,117151		0,102288		0,111731		0,0448		-0,2558	
	0,237078		0,2351		0,23625		0,236481		0,235251		0,236834		0,23625		0,251313	
d2002	0,238408		0,249004		0,232213		0,23448		0,222455		0,238458		0,232213		-0,03382	
	0,249183		0,250809		0,252166		0,247368		0,2469		0,248814		0,252166		0,257373	
d2003	0,194588		0,203357		0,152106		0,199117		0,186799		0,196009		0,152106		-0,15784	
	0,231621		0,232067		0,227236		0,226065		0,226452		0,225856		0,227236		0,247232	
d2004	0,128744		0,154301		0,074496		0,118189		0,107041		0,115747		0,074496		-0,10619	
	0,241319		0,242701		0,240136		0,233521		0,233618		0,234163		0,240136		0,247962	
d2005	0,215021		0,186331		0,204829		0,175968		0,172143		0,172695		0,204829		0,051691	
	0,31064		0,314264		0,312247		0,305487		0,305057		0,306843		0,312247		0,310424	
d01Laval	0,090771		0,120866		0,02597		-0,04514		-0,05164		-0,05113		0,02597		-0,01121	
	0,153336		0,153773		0,161384		0,16262		0,163497		0,164177		0,161384		0,159102	
d04MtlG	0,044029		0,072514		-0,03281		-0,1053		-0,10738		-0,11302		-0,03281		-0,06175	
	0,148968		0,147051		0,158698		0,156991		0,156705		0,160176		0,158698		0,156622	
d07Concordia	0,117489		0,136761		0,05734		0,118201		0,123983		0,102216		0,05734		0,336181	
	0,238731		0,236191		0,251284		0,256065		0,254559		0,258848		0,251284		0,252039	
d09UQAMG	0,118209		0,146516		0,047139		0,034664		0,036092		0,026728		0,047139		-0,15533	
	0,216931		0,216557		0,222948		0,21432		0,214394		0,215126		0,222948		0,220862	
d08SherbrookeG	0,315265		0,359711	*	0,201283		0,160098		0,158862		0,151717		0,201283		0,307809	
	0,215711		0,214001		0,223584		0,224489		0,223832		0,226262		0,223584		0,226287	
L_nbart_res			-0,04653													

nbreg de Bloc E – nombre de revendications avec sous-catégories de brevets																
Variable	nbreg_vce		nbreg_art		nbreg_vce1		nbreg_subv		nbreg_subv1		nbreg_subv2		nbreg_vce2		nbreg_cont	
Endogénéité			x				x		x		x				x	
			0,030734													
L1.ltotsubvmoyan3					-0,01088		0,069038 *		0,068497 *		0,066575		-0,01088		-0,06845 **	
					0,019503		0,039265		0,038816		0,042784		0,019503		0,028048	
L_subv_res							-0,09046 *									
							0,046926									
L_subv_res1									-0,08933 *							
									0,046534							
L_subv_res2											-0,08711 *					
											0,048883					
L_cont_res															-0,10897 ***	
															0,034432	
_cons	1,749822 ***		1,985799 ***		1,551105 ***		1,030149		1,016293		1,049177		1,551105 ***		1,466016 **	
	0,575653		0,637245		0,581212		0,684832		0,631815		0,685295		0,581212		0,576536	
Inalpha																
_cons	-0,4776 ***		-0,48159 ***		-0,46317 ***		-0,46927 ***		-0,46942 ***		-0,46829 ***		-0,46317 ***		-0,48193 ***	
	0,070615		0,070832		0,073715		0,074036		0,073752		0,074533		0,073715		0,075324	
Statistics																
N	422		419		422		419		422		419		422		422	
df_m	34		35		34		35		35		35		34		35	
k	36		37		36		37		37		37		36		37	
k_aux	1		1		1		1		1		1		1		1	
k_eq	2		2		2		2		2		2		2		2	
k_eq_model	1		1		1		1		1		1		1		1	
k_dv	1		1		1		1		1		1		1		1	
ll	-1763,24		-1750,4		-1766,56		-1753,24		-1765,15		-1753,47		-1766,56		-1762,42	
ll_0	-1859,17		-1847,4		-1859,17		-1847,4		-1859,17		-1847,4		-1859,17		-1859,17	
chi2	267,955		296,0708		274,9542		288,8908		288,3335		286,7237		274,9542		283,0475	
aic	3598,487		3574,796		3605,128		3580,486		3604,299		3580,945		3605,128		3598,838	
bic	3744,107		3724,198		3750,748		3729,887		3753,964		3730,346		3750,748		3748,503	
p	3,81E-38		4,35E-43		1,73E-39		1,05E-41		1,35E-41		2,75E-41		1,73E-39		1,4E-40	

Tableau R-4: Régressions à Moindres Carrés Ordinaires (deuxième étape) du bloc E – nombre de revendications avec sous-catégories de brevets

MCO de Bloc E – nombre de revendications avec sous-catégories de brevets															
Variable	reg_vce		reg_art		reg_vce1		reg_subv		reg_subv1		reg_subv2		reg_vce2		reg_cont
Endogénéité			x				x		x		x				x
L1.nbarticlemoyan3	0,032668	**	0,048489	*											
	0,016261		0,025448												
L1.ltotcontmoyan3	-0,07984	**	-0,07701	*	-0,08906	**	-0,09422	**	-0,09186	**	-0,0943	**	-0,08906	**	0,031164
	0,039626		0,040209		0,040703		0,040517		0,04071		0,040684		0,040703		0,079896
L1.sq_ltotcontmoyan3	0,007826	**	0,007564	**	0,008797	***	0,008615	***	0,008329	**	0,008602	***	0,008797	***	0,005593
	0,003266		0,003345		0,003333		0,003299		0,003331		0,003307		0,003333		0,004019
L1.lmliquind23	-0,03314		-0,06777		0,042314		0,048701		0,044291		0,048764		0,042314		0,052861
	0,124143		0,138068		0,117908		0,118952		0,118021		0,119015		0,117908		0,115287
L1.sq_lmliquind23	0,002173		0,003594		-0,00093		-0,00129		-0,00111		-0,0013		-0,00093		-0,00087
	0,004877		0,005456		0,004607		0,004649		0,004613		0,004651		0,004607		0,004512
deg_app	-0,06218		-0,0646		-0,06878		-0,09057		-0,08566		-0,09098		-0,06878		-0,06061
	0,079313		0,081515		0,079559		0,081138		0,079906		0,081416		0,079559		0,079818
nbloop	0,175896	***	0,178274	***	0,172732	***	0,16152	***	0,160025	***	0,161167	***	0,172732	***	0,183687
	0,028778		0,02944		0,028911		0,031401		0,03149		0,030869		0,028911		0,030313
propuninv	-0,38912	*	-0,41563	*	-0,39237	*	-0,47189	**	-0,43999	*	-0,47316	**	-0,39237	*	-0,40105
	0,223054		0,221328		0,227144		0,228148		0,229145		0,232043		0,227144		0,22523
metrology_processus_pat	0,197477		0,186761		0,216355		0,215624		0,219623		0,215717		0,216355		0,348716
	0,23135		0,228498		0,222175		0,22408		0,225415		0,223752		0,222175		0,221933
devices_electronics_pat	0,40767	**	0,394402	**	0,425796	**	0,385107	**	0,390343	**	0,384155	**	0,425796	**	0,587488
	0,184188		0,185881		0,181184		0,189029		0,188702		0,190306		0,181184		0,21083
structures_materials_pat	0,458553	**	0,461741	**	0,450599	**	0,495567	***	0,498784	***	0,496838	***	0,450599	**	0,20465
	0,193201		0,193042		0,184504		0,189359		0,190183		0,189938		0,184504		0,206694
dna_rna_pat	0,251335		0,250243		0,240496		0,224742		0,231613		0,224506		0,240496		0,2393
	0,161109		0,161715		0,161299		0,16216		0,161504		0,161951		0,161299		0,16019
proteins_molecules_pat	0,407125	**	0,41514	**	0,381185	**	0,416944	**	0,421663	**	0,418091	**	0,381185	**	0,404331
	0,173133		0,172965		0,171761		0,178875		0,17798		0,178744		0,171761		0,172243
cell_tissue_engineerin_pat	0,896568	***	0,915428	***	0,851682	***	0,877783	***	0,881773	***	0,87862	***	0,851682	***	0,71824
	0,176564		0,18016		0,174661		0,174537		0,174441		0,176611		0,174661		0,183747
process_biotech_pat	0,647073	***	0,693964	***	0,620427	***	0,690272	***	0,654069	***	0,691095	***	0,620427	***	0,454198
	0,172561		0,167103		0,164513		0,165623		0,168153		0,163216		0,164513		0,168006
gene_rna_vector_pat	0,445732	***	0,440647	***	0,444464	***	0,474181	***	0,482114	***	0,475205	***	0,444464	***	0,505353
	0,152822		0,15188		0,154576		0,156681		0,157033		0,158527		0,154576		0,159712
bioinformatics_pat	0,539652	**	0,521472	**	0,568156	**	0,573483	**	0,57873	**	0,573741	**	0,568156	**	0,740266
	0,258772		0,257377		0,272688		0,269456		0,26859		0,270377		0,272688		0,296255
codechair33	-0,04874		-0,0564		-0,03899		-0,05962		-0,05689		-0,06011		-0,03899		-0,03942

MCO de Bloc E – nombre de revendications avec sous-catégories de brevets								
Variable	reg_vce	reg_art	reg_vce1	reg_subv	reg_subv1	reg_subv2	reg_vce2	reg_cont
Endogénéité	x		x		x		x	
	0,061291	0,062162	0,06645	0,068638	0,068222	0,071425	0,06645	0,065683
age	0,030962	0,016733	0,056821	0,054842	0,054313	0,054653	0,056821	0,106791
	0,073368	0,085604	0,070068	0,080959	0,070629	0,080971	0,070068	0,077617
sq_age	-0,0017	-0,0013	-0,00239	-0,00255	-0,00253	-0,00255	-0,00239	-0,0045
	0,00288	0,003253	0,002817	0,003162	0,002833	0,003165	0,002817	0,003115
d1997	0,124078	0,116841	0,133549	0,243294	0,249426	0,246446	0,133549	0,033093
	0,225315	0,227191	0,22771	0,235988	0,238441	0,230827	0,22771	0,22216
d1998	0,181864	0,226832	0,173071	0,226667	0,185176	0,226983	0,173071	-0,01856
	0,218878	0,223879	0,221142	0,221568	0,217683	0,221171	0,221142	0,22573
d1999	0,157566	0,147864	0,183816	0,216713	0,209985	0,21758	0,183816	-0,04229
	0,232113	0,237307	0,23711	0,238258	0,235646	0,2385	0,23711	0,257697
d2000	0,216199	0,20531	0,251467	0,230098	0,220702	0,22933	0,251467	0,076102
	0,226906	0,231827	0,230665	0,233445	0,231944	0,236108	0,230665	0,236772
d2001	-0,04561	-0,05331	-0,02859	0,012102	0,009805	0,013248	-0,02859	-0,25314
	0,267769	0,271794	0,270396	0,27406	0,272378	0,272448	0,270396	0,284746
d2002	0,159438	0,153924	0,168136	0,159048	0,157195	0,158844	0,168136	-0,02766
	0,277023	0,280541	0,281847	0,28278	0,28218	0,284113	0,281847	0,281013
d2003	0,24388	0,253926	0,22908	0,252213	0,246409	0,252744	0,22908	0,029805
	0,268821	0,269032	0,268254	0,266852	0,266873	0,265998	0,268254	0,285175
d2004	0,227476	0,239097	0,207024	0,231393	0,226079	0,231959	0,207024	0,088679
	0,274742	0,27524	0,274504	0,272343	0,272691	0,271555	0,274504	0,275295
d2005	0,235669	0,216945	0,255287	0,228074	0,23434	0,227573	0,255287	0,147489
	0,392061	0,393439	0,393211	0,393682	0,393152	0,394947	0,393211	0,390728
d01Laval	0,157669	0,1858	0,119492	0,078057	0,063878	0,076529	0,119492	0,076327
	0,167372	0,169662	0,176403	0,182178	0,184329	0,181741	0,176403	0,176372
d04MtlG	0,087167	0,107975	0,043945	-0,003	-0,00522	-0,00427	0,043945	0,008038
	0,168736	0,169859	0,177429	0,183087	0,182114	0,1829	0,177429	0,175946
d07Concordia	0,282737	0,31654 *	0,21466	0,247249	0,248306	0,248137	0,21466	0,455494 **
	0,179643	0,183982	0,187376	0,189126	0,188776	0,19244	0,187376	0,205758
d09UQAMG	0,16375	0,17252	0,12248	0,102724	0,112901	0,102542	0,12248	-0,04538
	0,264342	0,265782	0,271245	0,268888	0,268674	0,269067	0,271245	0,281254
d08SherbrookeG	0,238328	0,268017	0,176322	0,143963	0,141407	0,143008	0,176322	0,246115
	0,268694	0,271467	0,277142	0,27865	0,279091	0,277817	0,277142	0,28714
L_nbart_res		-0,03186						
		0,032971						
L1.ltotsubvmoyan3			-0,00178	0,046857	0,049248	0,048243	-0,00178	-0,04896
			0,020854	0,05072	0,050257	0,056364	0,020854	0,032701

MCO de Bloc E – nombre de revendications avec sous-catégories de brevets																
Variable	reg_vce		reg_art		reg_vce1		reg_subv		reg_subv1		reg_subv2		reg_vce2		reg_cont	
Endogénéité			x				x		x		x				x	
L_subv_res							-0,05526									
							0,054116									
L_subv_res1									-0,05672							
									0,053295							
L_subv_res2											-0,05623					
											0,059205					
L_cont_res															-0,08806    **	
															0,043498	
_cons	2,220822	***	2,365768	***	2,016291	***	1,696808	**	1,657671	**	1,687635	**	2,016291	***	1,931603	***
	0,680771		0,74772		0,701816		0,831023		0,777249		0,848341		0,701816		0,706926	
	0,0013		0,0018		0,0044		0,0423		0,034		0,0478		0,0044		0,0068	
N	422		419		422		419		422		419		422		422	
df_m	34		35		34		35		35		35		34		35	
rank	35		36		35		36		36		36		35		36	
ll	-559,517		-555,402		-561,621		-557,301		-561,158		-557,322		-561,621		-559,758	
ll_0	-621,56		-618,321		-621,56		-618,321		-621,56		-618,321		-621,56		-621,56	
aic	1189,034		1182,804		1193,243		1186,601		1194,315		1186,643		1193,243		1191,517	
bic	1330,61		1328,167		1334,818		1331,965		1339,936		1332,007		1334,818		1337,137	
r2	0,254754		0,259427		0,247285		0,252685		0,248938		0,25261		0,247285		0,253902	



## ANNEXE S – Régressions de deuxième étape du bloc F

Tableau S-1: Régressions à binomiale négative (deuxième étape) du bloc F – nombre de citations

Nbreg du Bloc F – nombre de citations											
Variable	nbreg_vce	nbreg_art	nbreg_vce1	nbreg_subv	nbreg_subv1	nbreg_subv2	nbreg_vce2	nbreg_cont			
Endogénéité	x		x		x		x		x		
L1.nbarticlemoyan3	-0.0139 (0.0265)	-0.0270 (0.0360)									
L1.ltotcontmoyan3	-0.0753 (0.0624)	-0.0835 (0.0634)	-0.0614 (0.0628)	-0.0717 (0.0627)	-0.0697 (0.0627)	-0.0700 (0.0630)	-0.0614 (0.0628)	-0.0274 (0.1299)			
L1.sq_ltotcontmoyan3	0.0058 (0.0049)	0.0064 (0.0050)	0.0055 (0.0049)	0.0048 (0.0049)	0.0045 (0.0049)	0.0052 (0.0049)	0.0055 (0.0049)	0.0047 (0.0056)			
L1.lmcliquind23	0.3610 * (0.2163)	0.3997 * (0.2259)	0.3191 (0.1993)	0.3317 * (0.1986)	0.3236 (0.1978)	0.3334 * (0.1982)	0.3191 (0.1993)	0.3222 (0.1993)			
L1.sq_lmcliquind23	-0.0142 * (0.0085)	-0.0158 * (0.0088)	-0.0123 (0.0077)	-0.0131 * (0.0077)	-0.0128 * (0.0076)	-0.0131 * (0.0076)	-0.0123 (0.0077)	-0.0123 (0.0077)			
deg_app	0.1299 (0.1367)	0.1409 (0.1378)	0.1575 (0.1420)	0.1384 (0.1468)	0.1233 (0.1466)	0.1477 (0.1461)	0.1575 (0.1420)	0.1602 (0.1407)			
nbloop	0.1780 *** (0.0234)	0.1746 *** (0.0236)	0.1945 *** (0.0279)	0.1716 *** (0.0352)	0.1717 *** (0.0354)	0.1789 *** (0.0359)	0.1945 *** (0.0279)	0.1975 *** (0.0323)			
propuninv	-0.4794 (0.3218)	-0.4935 (0.3235)	-0.4188 (0.3322)	-0.5496 (0.3374)	-0.5509 (0.3366)	-0.5011 (0.3340)	-0.4188 (0.3322)	-0.4205 (0.3328)			
excl_nano_pat	1.0147 *** (0.2529)	1.0090 *** (0.2556)	1.0144 *** (0.2540)	1.0645 *** (0.2573)	1.0789 *** (0.2569)	1.0463 *** (0.2563)	1.0144 *** (0.2540)	1.0290 *** (0.2590)			
excl_bio_pat	-0.5093 * (0.2852)	-0.5111 * (0.2876)	-0.5466 * (0.2802)	-0.4525 (0.2837)	-0.4531 (0.2821)	-0.4780 * (0.2862)	-0.5466 * (0.2802)	-0.5654 ** (0.2876)			
codechair33	0.1751 * (0.1041)	0.1749 (0.1071)	0.1992 * (0.1051)	0.1576 (0.1084)	0.1578 (0.1084)	0.1732 (0.1085)	0.1992 * (0.1051)	0.1962 * (0.1048)			
age	0.1168 (0.1486)	0.1069 (0.1836)	0.1379 (0.1336)	0.0961 (0.1666)	0.1247 (0.1321)	0.0992 (0.1639)	0.1379 (0.1336)	0.1559 (0.1425)			
sq_age	-0.0035 (0.0059)	-0.0032 (0.0070)	-0.0042 (0.0055)	-0.0032 (0.0066)	-0.0042 (0.0054)	-0.0031 (0.0066)	-0.0042 (0.0055)	-0.0049 (0.0058)			
d1997	0.0340 (0.4025)	0.0321 (0.3964)	-0.1720 (0.3509)	0.0771 (0.3911)	0.0963 (0.3947)	0.0016 (0.4020)	-0.1720 (0.3509)	-0.1966 (0.3695)			
d1998	-0.6517 ** (0.2789)	-0.6463 ** (0.2883)	-0.6156 ** (0.2830)	-0.5756 * (0.3002)	-0.5823 ** (0.2898)	-0.5816 * (0.2998)	-0.6156 ** (0.2830)	-0.6648 ** (0.3359)			
d1999	-0.6285 * (0.3387)	-0.5742 * (0.3419)	-0.6023 * (0.3409)	-0.4744 (0.3546)	-0.5215 (0.3508)	-0.4954 (0.3585)	-0.6023 * (0.3409)	-0.6752 (0.4612)			
d2000	-0.5923 * (0.3552)	-0.5582 (0.3563)	-0.5419 (0.3566)	-0.5601 (0.3569)	-0.5964 (0.3585)	-0.5374 (0.3550)	-0.5419 (0.3566)	-0.5949 (0.4109)			
d2001	-0.6865 * (0.4119)	-0.6507 (0.4107)	-0.7238 * (0.3929)	-0.5924 (0.4054)	-0.6268 (0.4054)	-0.6137 (0.4099)	-0.7238 * (0.3929)	-0.7913 (0.4932)			

Nbreg du Bloc F – nombre de citations																
Variable	nbreg_vce		nbreg_art		nbreg_vce1		nbreg_subv		nbreg_subv1		nbreg_subv2		nbreg_vce2		nbreg_cont	
Endogénéité			x				x		x		x				x	
d2002	-1.3708	***	-1.3321	***	-1.3113	***	-1.2649	**	-1.3027	***	-1.2601	**	-1.3113	***	-1.3850	**
	(0.4885)		(0.4881)		(0.4925)		(0.4955)		(0.4965)		(0.4936)		(0.4925)		(0.5702)	
d2003	-2.2960	***	-2.2861	***	-2.3031	***	-2.2544	***	-2.2884	***	-2.2483	***	-2.3031	***	-2.3638	***
	(0.5388)		(0.5350)		(0.5414)		(0.5402)		(0.5375)		(0.5441)		(0.5414)		(0.6035)	
d2004	-2.7911	***	-2.7645	***	-2.7488	***	-2.6814	***	-2.7048	***	-2.6878	***	-2.7488	***	-2.7851	***
	(0.6451)		(0.6457)		(0.6495)		(0.6579)		(0.6552)		(0.6576)		(0.6495)		(0.6628)	
d2005	-2.9264	***	-2.8926	***	-2.8710	***	-2.8654	***	-2.8960	***	-2.8501	***	-2.8710	***	-2.9140	***
	(0.6662)		(0.6652)		(0.6768)		(0.6741)		(0.6759)		(0.6740)		(0.6768)		(0.7066)	
d01Laval	0.1012		0.0905		0.2646		0.1274		0.1110		0.1757		0.2646		0.2500	
	(0.2807)		(0.2796)		(0.2587)		(0.2702)		(0.2733)		(0.2613)		(0.2587)		(0.2705)	
d04MtlG	0.6542	**	0.6462	**	0.7837	***	0.6851	***	0.6580	**	0.7207	***	0.7837	***	0.7698	***
	(0.2888)		(0.2794)		(0.2684)		(0.2604)		(0.2639)		(0.2576)		(0.2684)		(0.2808)	
d07Concordia	0.0638		0.0526		0.2145		0.2730		0.2840		0.2445		0.2145		0.2889	
	(0.4409)		(0.4439)		(0.4406)		(0.4478)		(0.4492)		(0.4446)		(0.4406)		(0.4811)	
d09UQAMG	0.8392	***	0.8334	***	0.9278	***	0.8636	***	0.8389	***	0.8911	***	0.9278	***	0.8577	**
	(0.3001)		(0.3004)		(0.2918)		(0.2856)		(0.2890)		(0.2849)		(0.2918)		(0.3950)	
d08SherbrookeG	-0.0162		-0.0402		0.0974		-0.0106		-0.0237		0.0276		0.0974		0.1090	
	(0.4353)		(0.4332)		(0.4156)		(0.4349)		(0.4300)		(0.4313)		(0.4156)		(0.4151)	
L_nbart_res			0.0224													
			(0.0526)													
L1.ltotsubvmoyan3					-0.0655	**	0.0415		0.0514		0.0046		-0.0655	**	-0.0797	
					(0.0326)		(0.0904)		(0.0913)		(0.0927)		(0.0326)		(0.0590)	
L_subv_res							-0.1266									
							0.0954									
L_subv_res1									-0.1399							
									(0.0985)							
L_subv_res2											-0.0800					
											(0.0932)					
L_cont_res															-0.0268	
															(0.0886)	
_cons	-0.7936		-0.8052		-0.4942		-1.1007		-1.2912		-0.8637		-0.4942		-0.5280	
	(1.1661)		(1.3712)		(1.1764)		(1.4803)		(1.2676)		(1.5549)		(1.1764)		(1.1579)	
Inalpha																
_cons	0.6490		0.6563		0.6334		0.6386		0.6295		0.6402		0.6334		0.6331	
	(0.1312)		(0.1304)		(0.1282)		(0.1283)		(0.1286)		(0.1283)		(0.1282)		(0.1284)	
Statistics																
N	422		419		422		419		422		419		422		422	
df_m	27		28		27		28		28		28		27		28	
k	29		30		29		30		30		30		29		30	
k_aux	1		1		1		1		1		1		1		1	
k_eq	2		2		2		2		2		2		2		2	
k_eq_model	1		1		1		1		1		1		1		1	

Nbreg du Bloc F – nombre de citations								
Variable	nbreg_vce	nbreg_art	nbreg_vce1	nbreg_subv	nbreg_subv1	nbreg_subv2	nbreg_vce2	nbreg_cont
Endogénéité	x	x	x	x	x	x	x	x
k_dv	1	1	1	1	1	1	1	1
ll	-706.99393	-700.95705	-705.16378	-698.67404	-704.45184	-699.07376	-705.16378	-705.12739
ll_0	-769.19194	-762.78897	-769.19194	-762.78897	-769.19194	-762.78897	-769.19194	-769.19194
chi2	222.55783	222.56652	209.33904	240.87889	237.77832	251.00392	209.33904	222.85112
aic	1471.9879	1461.9141	1468.3276	1457.3481	1468.9037	1458.1475	1468.3276	1470.2548
bic	1589.293	1583.0502	1585.6327	1578.4842	1590.2538	1579.2836	1585.6327	1591.6049
p	1.18E-32	3.41E-32	4.09E-30	9.97E-36	3.98E-35	1.07E-37	4.09E-30	3.01E-32

Tableau S-2: Régressions tobit sans endogénéité et de deuxième étape du bloc F – nombre de citations

Tobit du Bloc F – nombre de citations					
Variable	HSM_tobit_art	HSM_tobit_subv	HSM_tobit_subv_1	HSM_tobit_subv_2	HSM_tobit_cont
Endogénéité	x	x	x	x	x
L_nbarticlemoyan3	0.0146 (0.0357)				
L_itotcontmoyan3	-0.0057 (0.0600)	-0.0140 (0.0597)	-0.0097 (0.0593)	-0.0131 (0.0597)	0.0485 (0.1219)
L_sq_itotcontmoyan3	0.0000 (0.0050)	-0.0006 (0.0050)	-0.0008 (0.0050)	-0.0004 (0.0049)	-0.0007 (0.0055)
L_lmcliqind23	0.1980 (0.2270)	0.2379 (0.1972)	0.2294 (0.1952)	0.2404 (0.1973)	0.2298 (0.1964)
L_sq_lmcliqind23	-0.0072 (0.0089)	-0.0090 (0.0076)	-0.0087 (0.0076)	-0.0091 (0.0077)	-0.0082 (0.0076)
deg_app	-0.0342 (0.1243)	-0.0670 (0.1224)	-0.0835 (0.1212)	-0.0611 (0.1226)	-0.0315 (0.1225)
nbloop	0.1959 *** (0.0338)	0.1731 *** (0.0400)	0.1753 *** (0.0403)	0.1776 *** (0.0416)	0.2067 *** (0.0393)
propuninv	-0.2390 (0.3059)	-0.3335 (0.3027)	-0.3322 (0.2977)	-0.3091 (0.3084)	-0.1924 (0.3003)
excl_nano_pat	0.8829 *** (0.2451)	0.9101 *** (0.2474)	0.9220 *** (0.2476)	0.9032 *** (0.2455)	0.9004 *** (0.2516)
excl_bio_pat	-0.1216 (0.2355)	-0.0710 (0.2369)	-0.0773 (0.2359)	-0.0893 (0.2345)	-0.2058 (0.2326)
codechair33	0.0584 (0.0956)	0.0300 (0.0961)	0.0335 (0.0955)	0.0387 (0.0989)	0.0776 (0.0950)
age	0.0411 (0.1417)	0.0410 (0.1371)	0.0734 (0.1183)	0.0435 (0.1368)	0.1073 (0.1269)
sq_age	-0.0023 (0.0055)	-0.0027 (0.0054)	-0.0037 (0.0048)	-0.0026 (0.0055)	-0.0044 (0.0052)
d1997	-0.1378 (0.3476)	0.1010 (0.3859)	0.0962 (0.3868)	0.0500 (0.3949)	-0.2541 (0.3463)
d1998	-0.1094 (0.2872)	-0.0953 (0.2784)	-0.0929 (0.2707)	-0.0993 (0.2786)	-0.1740 (0.2839)
d1999	-0.2602 (0.3244)	-0.1772 (0.3220)	-0.2428 (0.3179)	-0.1887 (0.3225)	-0.3999 (0.3745)
d2000	-0.3375 (0.3410)	-0.3656 (0.3318)	-0.3985 (0.3317)	-0.3516 (0.3321)	-0.3740 (0.3558)
d2001	-0.6195 (0.3868)	-0.5229 (0.3874)	-0.5638 (0.3846)	-0.5388 (0.3884)	-0.7511 * (0.4129)
d2002	-1.0054 ** (0.4441)	-1.0020 ** (0.4384)	-1.0421 ** (0.4370)	-0.9987 ** (0.4384)	-1.1085 ** (0.4731)
d2003	-1.6630 *** (0.4664)	-1.6454 *** (0.4515)	-1.6786 *** (0.4506)	-1.6488 *** (0.4524)	-1.7756 *** (0.4732)
d2004	-1.8650 *** (0.5156)	-1.8242 *** (0.5093)	-1.8477 *** (0.5063)	-1.8333 *** (0.5108)	-1.9272 *** (0.5091)
d2005	-1.6841 *** (0.5699)	-1.6941 *** (0.5635)	-1.7239 *** (0.5635)	-1.6890 *** (0.5630)	-1.7245 *** (0.5761)
d01Laval	0.1258 (0.2230)	-0.0045 (0.2491)	-0.0010 (0.2482)	0.0235 (0.2492)	0.1580 (0.2279)
d04MtlG	0.4442 * (0.2275)	0.3384 (0.2313)	0.3191 (0.2327)	0.3561 (0.2332)	0.4302 * (0.2340)
d07Concordia	0.5239 ** (0.2473)	0.5734 ** (0.2416)	0.5789 ** (0.2409)	0.5578 ** (0.2421)	0.6028 * (0.3250)
d09UQAMG	0.8651 *** (0.2652)	0.8230 *** (0.2519)	0.8009 *** (0.2514)	0.8295 *** (0.2543)	0.7530 ** (0.3517)
d08SherbrookeG	-0.0817 (0.3545)	-0.2096 (0.3630)	-0.2072 (0.3607)	-0.1927 (0.3631)	-0.0620 (0.3501)
L_nbart_res	0.0084 (0.0488)				
L_itotsubvmoyan3		0.0996 (0.0790)	0.0991 (0.0788)	0.0768 (0.0883)	-0.0635 (0.0535)
L_subv_res		-0.1550 * 0.0833			
L_subv_res1			-0.1569 * (0.0835)		
L_subv_res2				-0.1277	

Tobit du Bloc F – nombre de citations					
Variable	HSM_tobit_art	HSM_tobit_subv	HSM_tobit_subv_1	HSM_tobit_subv_2	HSM_tobit_cont
Endogénéité	x	x	x	x	x
L_cont_res				(0.0908)	-0.0389 (0.0808)
_cons	-0.1214 (1.1153)	-0.8942 (1.2110)	-1.0207 (1.1116)	-0.7482 (1.2885)	-0.0929 (1.0097)
sigma					
_cons	1.2845478 0.06492178	1.2734219 0.06473247	1.2689022 0.06451303	1.2753608 0.06480035	1.2757662 0.06406704
Statistics					
N	419	419	422	419	422
df_m	28	28	28	28	28
k_aux	1	1	1	1	1
rank	30	30	30	30	30
ll	-466.04052	-463.88347	-467.26586	-464.43306	-468.58964
ll_0	-529.28907	-529.28907	-533.47388	-529.28907	-533.47388
aic	992.08103	987.76694	994.53173	988.86612	997.17928
bic	1113.2172	1108.9031	1115.8819	1110.0022	1118.5294
r2					
r2_p	0.11949718	0.12357255	0.12410733	0.12253419	0.1216259
p	2.62E-18	4.91E-22	9.23E-22	2.24E-22	6.60E-19

Tableau S-3: Régressions à binomiale négative (deuxième étape) du bloc F – nombre de citations avec sous-catégories de brevets

nbreg de Bloc F nombre de citations avec sous-catégories de brevets												
Variable	nbreg_vce	nbreg_art	nbreg_vce1	nbreg_subv	nbreg_subv1	nbreg_subv2	nbreg_vce2	nbreg_cont				
Endogénéité	x		x		x		x		x			
L1.nbarticlemoyan3	-0,00678	-0,03388										
	0,020288	0,037743										
L1.ltotcontmoyan3	-0,08016	-0,09391	-0,07417	-0,07916	-0,07749	-0,07692	-0,07417	0,04575				
	0,061382	0,063771	0,061384	0,061676	0,06174	0,061931	0,061384	0,145562				
L1.sq_ltotcontmoyan3	0,007113	0,008377	0,007338	0,006957	0,006677	0,00738	0,007338	0,004801				
	0,005093	0,005306	0,005067	0,005035	0,005039	0,005017	0,005067	0,00574				
L1.lmliquind23	0,207315	0,276341	0,187615	0,194348	0,184801	0,196598	0,187615	0,201517				
	0,205638	0,221111	0,197187	0,199	0,19739	0,198362	0,197187	0,197888				
L1.sq_lmliquind23	-0,00792	-0,01078	-0,007	-0,00744	-0,00704	-0,00743	-0,007	-0,0071				
	0,008041	0,008664	0,007641	0,00767	0,007617	0,007653	0,007641	0,007642				
deg_app	0,149994	0,162067	0,161034	0,15621	0,142169	0,166457	0,161034	0,158377				
	0,140144	0,141895	0,142057	0,144478	0,143472	0,144914	0,142057	0,142576				
nbloop	0,162469 ***	0,15756 ***	0,178703 ***	0,1618 ***	0,16185 ***	0,172699 ***	0,178703 ***	0,192253 ***				
	0,025902	0,02643	0,026075	0,034749	0,034452	0,035246	0,026075	0,032692				
propuninv	-0,43037	-0,44942	-0,40178	-0,47274	-0,46675	-0,43221	-0,40178	-0,39795				
	0,318162	0,317181	0,327807	0,330129	0,329124	0,327858	0,327807	0,328403				
metrology_processus_pat	0,838984 **	0,82561 **	0,851982 **	0,845896 **	0,851668 **	0,845926 **	0,851982 **	0,986138 ***				
	0,355802	0,355512	0,367283	0,368041	0,367253	0,367994	0,367283	0,377721				
devices_electronics_pat	0,597059 *	0,625491 *	0,636607 *	0,622177 *	0,609505 *	0,639199 *	0,636607 *	0,81737 *				
	0,352331	0,355899	0,361036	0,373383	0,369316	0,374726	0,361036	0,420968				
structures_materials_pat	0,261054	0,241751	0,19118	0,242307	0,258528	0,200909	0,19118	-0,07509				
	0,346585	0,349606	0,351275	0,359434	0,35953	0,364833	0,351275	0,450242				
dna_rna_pat	-0,32613	-0,33893	-0,29707	-0,31115	-0,30191	-0,30818	-0,29707	-0,31958				
	0,208211	0,208124	0,204414	0,204575	0,205157	0,204185	0,204414	0,209139				
proteins_molecules_pat	0,179554	0,162105	0,126882	0,184435	0,191761	0,141136	0,126882	0,138856				
	0,263586	0,264697	0,249597	0,253685	0,253164	0,251136	0,249597	0,246307				
cell_tissue_engineerin_pat	0,424329 *	0,403448	0,418752 *	0,451078 *	0,450126 *	0,430929 *	0,418752 *	0,267324				
	0,241413	0,248534	0,237343	0,237817	0,237577	0,238044	0,237343	0,285059				
process_biotech_pat	0,528741 *	0,518473 *	0,475785 *	0,519514 *	0,524127 *	0,491889 *	0,475785 *	0,310512				
	0,277617	0,282899	0,270754	0,275395	0,271748	0,276604	0,270754	0,316914				
gene_rna_vector_pat	0,321636	0,314982	0,276898	0,308706	0,326279	0,277669	0,276898	0,332797				
	0,233755	0,233341	0,226836	0,223577	0,225122	0,221216	0,226836	0,239145				

nbreg de Bloc F nombre de citations avec sous-catégories de brevets																
Variable	nbreg_vce		nbreg_art		nbreg_vce1		nbreg_subv		nbreg_subv1		nbreg_subv2		nbreg_vce2		nbreg_cont	
Endogénéité			x				x		x		x				x	
bioinformatics_pat	0,761336	*	0,785549	**	0,810103	**	0,795914	**	0,806142	**	0,801617	**	0,810103	**	0,996775	**
	0,392739		0,393527		0,397898		0,392573		0,388642		0,398938		0,397898		0,441074	
codechair33	0,093342		0,095773		0,110789		0,088645		0,089616		0,102122		0,110789		0,104299	
	0,104262		0,10722		0,105355		0,10649		0,106475		0,107799		0,105355		0,104569	
age	0,063985		0,055665		0,071831		0,020896		0,064544		0,02959		0,071831		0,124689	
	0,126395		0,153828		0,119404		0,143471		0,118333		0,141411		0,119404		0,125165	
sq_age	-0,00189		-0,00151		-0,00208		-0,0005		-0,00207		-0,00065		-0,00208		-0,00419	
	0,005034		0,005888		0,004893		0,005734		0,004852		0,005689		0,004893		0,005114	
d1997	0,17417		0,183169		0,013171		0,13592		0,168361		0,036575		0,013171		-0,10484	
	0,341242		0,341224		0,323166		0,362578		0,363506		0,370506		0,323166		0,345034	
d1998	-0,38368		-0,35189		-0,37324		-0,31666		-0,3534		-0,32973		-0,37324		-0,58531	*
	0,291765		0,301362		0,295917		0,307085		0,296756		0,308612		0,295917		0,355828	
d1999	-0,28383		-0,20858		-0,27585		-0,18723		-0,23444		-0,2187		-0,27585		-0,53987	
	0,334585		0,336843		0,337763		0,341935		0,338275		0,34457		0,337763		0,467197	
d2000	-0,38361		-0,33107		-0,36093		-0,34592		-0,3835		-0,3346		-0,36093		-0,57255	
	0,345137		0,34731		0,346804		0,345884		0,345706		0,347791		0,346804		0,405687	
d2001	-0,27799		-0,24102		-0,32253		-0,24614		-0,27366		-0,28127		-0,32253		-0,5765	
	0,46123		0,456688		0,450156		0,458455		0,457651		0,46047		0,450156		0,55588	
d2002	-0,89888	*	-0,84795	*	-0,86498	*	-0,83558	*	-0,8718	*	-0,83428	*	-0,86498	*	-1,10971	**
	0,493605		0,494176		0,497933		0,497829		0,497847		0,498233		0,497933		0,565684	
d2003	-1,84423	***	-1,85365	***	-1,88766	***	-1,85281	***	-1,8756	***	-1,86248	***	-1,88766	***	-2,13416	***
	0,527348		0,527023		0,5207		0,519035		0,516509		0,522798		0,5207		0,590319	
d2004	-2,43228	***	-2,40989	***	-2,43343	***	-2,38567	***	-2,40602	***	-2,40446	***	-2,43343	***	-2,61036	***
	0,629039		0,634367		0,628231		0,629852		0,624594		0,63203		0,628231		0,644097	
d2005	-2,83264	***	-2,79157	***	-2,80246	***	-2,79438	***	-2,80915	***	-2,79359	***	-2,80246	***	-2,94974	***
	0,637543		0,636534		0,651904		0,657293		0,655403		0,656019		0,651904		0,674837	
d01Laval	0,306191		0,278327		0,419911	*	0,355041		0,34064		0,405926	*	0,419911	*	0,376819	
	0,248617		0,249286		0,238525		0,244088		0,24326		0,245005		0,238525		0,248794	
d04MtlG	0,667726	**	0,637779	**	0,771689	***	0,731616	***	0,701962	***	0,773526	***	0,771689	***	0,744561	***
	0,285871		0,277946		0,280864		0,272277		0,270372		0,275414		0,280864		0,286883	
d07Concordia	0,050695		0,020586		0,165295		0,18899		0,197389		0,170514		0,165295		0,423358	
	0,327782		0,328293		0,308084		0,31809		0,32017		0,313652		0,308084		0,408973	
d09UQAMG	0,866768	***	0,842442	***	0,93574	***	0,914706	***	0,891954	***	0,941148	***	0,93574	***	0,756375	**
	0,300566		0,305012		0,295592		0,28991		0,288252		0,295501		0,295592		0,372276	
d08SherbrookeG	0,448064		0,393541		0,557677		0,490216		0,498217		0,526885		0,557677		0,639556	
	0,461927		0,457894		0,460794		0,470852		0,471369		0,469627		0,460794		0,464951	
L_nbart_res			0,046922													

nbreg de Bloc F nombre de citations avec sous-catégories de brevets																
Variable	nbreg_vce		nbreg_art		nbreg_vce1		nbreg_subv		nbreg_subv1		nbreg_subv2		nbreg_vce2		nbreg_cont	
Endogénéité			x				x		x		x				x	
			0,05203													
L1.ltotsubvmoyan3					-0,05079 *		0,002577		0,014337		-0,03858		-0,05079 *		-0,09954	
					0,028574		0,079154		0,079182		0,083194		0,028574		0,063916	
L_subv_res							-0,06205									
							0,087583									
L_subv_res1									-0,07745							
									0,088389							
L_subv_res2											-0,01261					
											0,087502					
L_cont_res															-0,09585	
															0,108956	
_cons	-1,13416	-1,18446	-0,80728	-0,9275	-1,23986	-0,67556	-0,80728	-0,89305								
	1,031051	1,213156	1,071403	1,338873	1,201534	1,367835	1,071403	1,051133								
Inalpha																
_cons	0,605569 ***	0,60822 ***	0,597473 ***	0,604059 ***	0,596449 ***	0,604621 ***	0,597473 ***	0,59461 ***								
	0,153784	0,152039	0,15257	0,151433	0,152456	0,151424	0,15257	0,153981								
Statistics																
N	422	419	422	419	422	419	422	422								
df_m	34	35	34	35	35	35	35	35								
k	36	37	36	37	37	37	37	37								
k_aux	1	1	1	1	1	1	1	1								
k_eq	2	2	2	2	2	2	2	2								
k_eq_model	1	1	1	1	1	1	1	1								
k_dv	1	1	1	1	1	1	1	1								
ll	-702,966	-696,6	-701,927	-695,753	-701,692	-695,901	-701,927	-701,481								
ll_0	-769,192	-762,789	-769,192	-762,789	-769,192	-762,789	-769,192	-769,192								
chi2	357,5276	324,4177	322,0325	323,8634	328,6977	340,1459	322,0325	341,4705								
aic	1477,933	1467,199	1475,853	1465,506	1477,383	1465,802	1475,853	1476,961								
bic	1623,553	1616,6	1621,473	1614,908	1627,048	1615,204	1621,473	1626,626								
p	1,32E-55	1,36E-48	1,28E-48	1,74E-48	1,98E-49	1,14E-51	1,28E-48	6,24E-52								



Tableau S-4: Régressions Tobit (deuxième étape) du bloc F - nombre de citations avec sous-catégories de brevets

Tobit de Bloc F - nombre de citations avec sous-catégories de brevets					
Variable	HSM_tobit_art	HSM_tobit_subv	HSM_tobit_subv_1	HSM_tobit_subv_2	HSM_tobit_cont
Endogénéité	x	x	x	x	x
L_itotcontmoyan3	-0,01542	-0,02398	-0,02003	-0,02286	0,081511
	0,058323	0,057832	0,057666	0,057883	0,126495
L_sq_itotcontmoyan3	0,000539	0,000229	-6,4E-05	0,000587	-0,00101
	0,004977	0,004977	0,004963	0,004918	0,005611
L_lmcliqind23	0,113049	0,160957	0,151338	0,163315	0,165942
	0,212565	0,182819	0,180441	0,183405	0,180924
L_sq_lmcliqind23	-0,00355	-0,00566	-0,00526	-0,00569	-0,00519
	0,008353	0,007104	0,007016	0,007133	0,007067
deg_app	-0,05403	-0,08586	-0,10117	-0,07748	-0,05855
	0,119445	0,118172	0,11655	0,118685	0,116944
nbloop	0,180719 ***	0,157723 ***	0,158538 ***	0,16584 ***	0,19699 ***
	0,032993	0,04112	0,040975	0,042159	0,039531
propuninv	-0,19463	-0,27083	-0,26254	-0,24199	-0,15491
	0,278283	0,274667	0,270431	0,277288	0,275955
metrology_processus_pat	0,603094	0,62842	0,63349 *	0,626041	0,738573 *
	0,379036	0,385634	0,383929	0,385815	0,387272
devices_electronics_pat	0,521433	0,465549	0,464644	0,484808	0,679871
	0,370514	0,394231	0,39343	0,394471	0,421298
structures_materials_pat	0,163504	0,232866	0,246583	0,205915	-0,07189
	0,384016	0,387258	0,388338	0,388568	0,445461
dna_rna_pat	-0,08757	-0,10459	-0,10069	-0,10158	-0,09398
	0,220708	0,218346	0,219755	0,218974	0,218732
proteins_molecules_pat	0,105622	0,149201	0,136187	0,122593	0,063868
	0,240938	0,247124	0,246668	0,246044	0,238445
cell_tissue_engineerin_pat	0,568779 **	0,586749 **	0,584517 **	0,567068 **	0,399178
	0,248116	0,23999	0,239781	0,240197	0,267982
process_biotech_pat	0,740906 ***	0,772125 ***	0,782544 ***	0,754118 ***	0,570812 *
	0,264222	0,259989	0,255597	0,25912	0,292802
gene_rna_vector_pat	0,265155	0,322072	0,329436	0,297301	0,292397
	0,215446	0,222122	0,22303	0,218513	0,223225
bioinformatics_pat	0,653304 **	0,704346 **	0,717231 **	0,701 **	0,84594 **
	0,319576	0,331044	0,32926	0,33814	0,375031
L_nbarticlemoyan3		0,01845			
		0,035746			
codechair33	0,007246	-0,01563	-0,01398	-0,00415	0,025411
	0,096526	0,097749	0,097029	0,099981	0,094664
age	0,012658	0,021261	0,055352	0,025264	0,101798
	0,132761	0,130439	0,110602	0,130427	0,116206
sq_age	-0,00156	-0,00211	-0,00328	-0,0021	-0,00467
	0,005195	0,005231	0,004544	0,005263	0,004843
L_nbart_res		0,01591			
		0,049066			
d1997	-0,03285	0,163256	0,171338	0,091487	-0,18105
	0,31469	0,341947	0,342201	0,347708	0,325097
d1998	0,037283	0,047507	0,03208	0,040808	-0,15579
	0,273109	0,270019	0,262131	0,272001	0,286113
d1999	-0,05195	0,022739	-0,03214	0,004845	-0,2825
	0,300651	0,299741	0,29656	0,300409	0,366031
d2000	-0,18898	-0,20517	-0,23609	-0,18663	-0,31777
	0,310575	0,309737	0,308579	0,311638	0,335685
d2001	-0,56319	-0,48527	-0,52177	-0,51038	-0,80215 **
	0,364533	0,36753	0,364958	0,368116	0,397758
d2002	-0,77422 *	-0,77978 *	-0,81426 *	-0,77279 *	-0,95573 **

Tobit de Bloc F - nombre de citations avec sous-catégories de brevets										
Variable	HSM_tobit_art		HSM_tobit_subv		HSM_tobit_subv_1		HSM_tobit_subv_2		HSM_tobit_cont	
Endogénéité	x		x		x		x		x	
	0,424711		0,422526		0,421217		0,42378		0,450889	
d2003	-1,38712	***	-1,37146	***	-1,39693	***	-1,38105	***	-1,60427	***
	0,443821		0,430104		0,429613		0,431999		0,458484	
d2004	-1,63243	***	-1,61228	***	-1,63264	***	-1,62625	***	-1,79441	***
	0,51238		0,506803		0,503605		0,509724		0,508369	
d2005	-1,60682	***	-1,6138	***	-1,63602	***	-1,6068	***	-1,69632	***
	0,526376		0,523754		0,522331		0,52428		0,534762	
d01Laval	0,189748		0,081182		0,073095		0,119339		0,180463	
	0,216712		0,229265		0,226755		0,230346		0,21673	
d04MtlG	0,352041		0,255596		0,231417		0,283379		0,3174	
	0,222055		0,219342		0,218253		0,222473		0,220234	
d07Concordia	0,504845		0,538311	*	0,550965	*	0,520545	*	0,696287	*
	0,314365		0,297969		0,30079		0,297765		0,376198	
d09UQAMG	0,780072	***	0,746779	***	0,723017	***	0,754731	***	0,621407	**
	0,267685		0,25065		0,247796		0,2549		0,310546	
d08SherbrookeG	0,006421		-0,10248		-0,10551		-0,07622		0,061785	
	0,364652		0,368476		0,367197		0,369033		0,365598	
L_ltotsubvmoyan3			0,082026		0,086366		0,050219		-0,07145	
			0,073966		0,072745		0,078802		0,054107	
L_subv_res			-0,12459							
			0,076675							
L_subv_res1					-0,1315	*				
					0,075716					
L_subv_res2							-0,08802			
							0,080161			
L_cont_res									-0,07547	
									0,083816	
_cons	-0,04188		-0,73792		-0,92072		-0,53424		-0,18534	
	1,026259		1,149886		1,04467		1,18554		0,91846	
sigma										
_cons	1,235649	***	1,230158	***	1,225428	***	1,232416	***	1,229981	***
	0,067735		0,067716		0,067241		0,067531		0,066861	
Statistics										
N	419		419		422		419		422	
df_m	35		35		35		35		35	
k_aux	1		1		1		1		1	
rank	37		37		37		37		37	
ll	-456,072		-455,172		-458,368		-455,725		-459,025	
ll_0	-529,289		-529,289		-533,474		-529,289		-533,474	
aic	986,1448		984,3447		990,7369		985,4507		992,0503	
bic	1135,546		1133,746		1140,402		1134,852		1141,716	
r2_p	0,13833		0,140031		0,140786		0,138986		0,139555	
p	6,3E-25		4,85E-29		1,14E-28		2,15E-29		2,82E-25	

## ANNEXE T – Régressions de première étape de phase 1 (blocs A, B et C)

Tableau T-1: Régressions de première étape en Moindres Carrés Ordinaires pour le bloc A – breveter et nombre de brevets

Variable	FS_reg_art		FS_reg_subv		FS_reg_subv1		FS_reg_subv2		FS_reg_cont	
L2.ltotsubvmoyan3	-0.1421 (0.0323)	***							-0.5283 (0.0810)	***
L2.sq_ltotsubvmoyan3	0.0125 (0.0029)	***							0.0653 (0.0072)	***
L2.ldmbetcent3	1.5588 (0.0731)	***	0.0762 (0.0671)				0.2188 (0.0487)	***	0.0617 (0.1358)	
codechair33	0.0000 (0.0327)		0.4362 (0.0457)	***	0.4374 (0.0457)	***	0.4391 (0.0460)	***	0.2429 (0.1379)	*
L1.age	-0.0494 (0.0229)	**	0.3241 (0.0470)	***	0.3287 (0.0470)	***	0.3206 (0.0471)	***	0.1615 (0.0771)	**
L1.sq_age	0.0022 (0.0011)	*	-0.0089 (0.0019)	***	-0.0091 (0.0019)	***	-0.0087 (0.0019)	***	-0.0072 (0.0035)	**
L1.nbarticlemoyan3wu	0.1605 (0.0784)	**								
L2.ltotcontmoyan3	0.0042 (0.0060)									
L1.ltotcontmoyan3			0.1242 (0.0071)	***	0.1244 (0.0071)	***	0.1247 (0.0072)	***		
L2.lmcliquind23	-0.0631 (0.0552)		-0.3834 (0.0764)	***	-0.3598 (0.0725)	***	-0.3941 (0.0760)	***	0.0264 (0.1675)	
L2.sq_lmcliquind23	0.0022 (0.0022)		0.0150 (0.0031)	***	0.0140 (0.0029)	***	0.0154 (0.0030)	***	-0.0007 (0.0066)	
L1.deg_app0	0.2309 (0.0085)	***	0.0601 (0.0205)	***	0.0624 (0.0206)	***	0.0661 (0.0204)	***	-0.1322 (0.0369)	***
L1.avg_pat_past	0.3068 (0.2199)		0.1974 (0.1193)	*	0.1944 (0.1153)	*	0.2334 (0.1445)			
L1.o_patpast	0.1226 (0.0676)	*	-0.0559 (0.0623)		-0.0529 (0.0620)		-0.0468 (0.0637)		0.0550 (0.1679)	
d1997	0.2339 (0.0630)	***	-2.9757 (0.1904)	***	-2.9813 (0.1903)	***	-2.9710 (0.1907)	***	-0.2155 (0.1616)	
d1998	0.1362 (0.0746)	*	-0.6436 (0.1419)	***	-0.6453 (0.1419)	***	-0.6314 (0.1423)	***	0.1808 (0.2147)	
d1999	0.2657 (0.0864)	***	-0.5529 (0.1452)	***	-0.5617 (0.1450)	***	-0.5318 (0.1453)	***	-0.1227 (0.2523)	
d2000	0.1310 (0.0873)		-0.9539 (0.1663)	***	-0.9611 (0.1661)	***	-0.9282 (0.1663)	***	0.0392 (0.2742)	
d2001	0.1540 (0.0912)	*	-1.2009 (0.1808)	***	-1.2041 (0.1806)	***	-1.1769 (0.1807)	***	-0.1627 (0.2923)	
d2002	0.2547 (0.0986)	***	-1.2481 (0.1929)	***	-1.2560 (0.1923)	***	-1.2186 (0.1925)	***	-0.3300 (0.3083)	
d2003	0.1931 (0.0991)	*	-1.5124 (0.2133)	***	-1.5173 (0.2129)	***	-1.4897 (0.2133)	***	-0.5998 (0.3215)	*
d2004	0.2649 (0.1012)	***	-1.6887 (0.2324)	***	-1.6979 (0.2316)	***	-1.6667 (0.2324)	***	-0.7712 (0.3273)	**
d2005	0.3053 (0.1076)	***	-1.7159 (0.2459)	***	-1.7269 (0.2450)	***	-1.6871 (0.2460)	***	-1.0092 (0.3400)	***
d01Laval	-0.0500 (0.1086)		0.0404 (0.1224)		0.0354 (0.1224)		0.0237 (0.1224)		1.6728 (0.3424)	***
d04MtlG	-0.0911 (0.1053)		0.1411 (0.1156)		0.1379 (0.1156)		0.1215 (0.1156)		1.3509 (0.3039)	***
d07Concordia	0.2845 (0.2074)		0.7626 (0.3066)	**	0.7572 (0.3055)	**	0.7597 (0.3074)	**	-0.3218 (0.3148)	
d09UQAMG	0.0425 (0.1366)		-0.1799 (0.1370)		-0.1908 (0.1363)		-0.1939 (0.1373)		2.0919 (0.3409)	***
d08SherbrookeG	-0.0605 (0.1173)		0.2816 (0.1498)	*	0.2774 (0.1499)	*	0.2622 (0.1496)	*	0.7504 (0.4036)	*
L2.nbarticlemoyan3			0.0819 (0.0262)	***	0.1047 (0.0191)	***				

Variable	FS_reg_art	FS_reg_subv	FS_reg_subv1	FS_reg_subv2	FS_reg_cont
L1.nbarticlemoyan3					0.0616 (0.0606)
L1.ltotsubvmoyan3wu		1.4058 *** (0.1416)	1.4047 *** (0.1416)	1.4085 *** (0.1417)	
L1.totcontmoyan3wu					0.1328 *** (0.0334)
L1.nb_loop_past					0.4845 *** (0.1366)
L1.totcont_past					0.0000 * (0.0000)
_cons	0.1164 (0.2402)	-7.6705 *** (1.5810)	-7.7153 *** (1.5771)	-7.6567 *** (1.5816)	-1.4852 ** (0.6238)
N	8772	8769	8769	8769	8771
df_m	27	26	25	25	28
rank	28	27	26	26	29
ll	-14942.5	-21345.7	-21346.6	-21351.5	-25609.4
ll_0	-19083.6	-22652	-22652	-22652	-26507.2
aic	29941.03	42745.46	42745.17	42754.94	51276.74
bic	30139.25	42936.6	42929.22	42938.99	51482.04
r2	0.610991	0.257645	0.257501	0.256673	0.185128

Tableau T-2: Régressions de première étape en Moindre Carrés Ordinaires en panel pour le bloc A – breveter et nombre de brevets

Variable	FS_xtreg_art		FS_xtreg_subv		FS_xtreg_subv1		FS_xtreg_subv2		FS_xtreg_cont	
L2.ltotsubvmoyan3	-0.0751 (0.0140)	***							-0.1970 (0.0464)	***
L2.sq_ltotsubvmoyan3	0.0064 (0.0012)	***							0.0285 (0.0040)	***
L2.ldmbetcent3	1.0788 (0.0197)	***	0.0170 (0.0635)				0.0587 (0.0492)		0.0744 (0.0748)	
codechair33	0.0346 (0.0325)		0.6365 (0.0771)	***	0.6368 (0.0770)	***	0.6379 (0.0771)	***	0.4114 (0.1124)	***
L1.age	0.0312 (0.0149)	**	0.5247 (0.0363)	***	0.5259 (0.0360)	***	0.5254 (0.0363)	***	0.2318 (0.0493)	***
L1.sq_age	-0.0003 (0.0006)		-0.0158 (0.0015)	***	-0.0158 (0.0015)	***	-0.0158 (0.0015)	***	-0.0071 (0.0020)	***
L1.nbarticlemoyan3wu	0.2655 (0.0652)	***								
L2.ltotcontmoyan3	0.0045 (0.0033)									
L1.ltotcontmoyan3			0.1076 (0.0080)	***	0.1077 (0.0080)	***	0.1077 (0.0080)	***		
L2.lmcliquind23	-0.0240 (0.0286)		-0.1204 (0.0716)	*	-0.1155 (0.0692)	*	-0.1223 (0.0716)	*	-0.0234 (0.0938)	
L2.sq_lmcliquind23	0.0003 (0.0011)		0.0046 (0.0028)		0.0044 (0.0027)		0.0046 (0.0028)		0.0010 (0.0037)	
L1.deg_app0	0.1974 (0.0072)	***	0.0339 (0.0181)	*	0.0340 (0.0181)	*	0.0347 (0.0181)	*	-0.0063 (0.0244)	
L1.avg_pat_past	-0.1246 (0.0721)	*	0.2708 (0.1773)		0.2716 (0.1772)		0.2720 (0.1773)		0.0346 (0.2634)	
L1.o_patpast	0.2717 (0.0315)	***	-0.1580 (0.0781)	**	-0.1578 (0.0781)	**	-0.1511 (0.0778)	*	0.0992 (0.1079)	
d1997	0.2368 (0.0591)	***	-2.8751 (0.1428)	***	-2.8764 (0.1427)	***	-2.8720 (0.1427)	***	-0.1712 (0.1893)	
d1998	0.1587 (0.0625)	**	-0.4176 (0.1455)	***	-0.4182 (0.1455)	***	-0.4107 (0.1453)	***	0.3701 (0.1880)	**
d1999	0.2360 (0.0677)	***	-0.4360 (0.1530)	***	-0.4379 (0.1529)	***	-0.4263 (0.1528)	***	0.2139 (0.1968)	
d2000	0.1052 (0.0679)		-0.8816 (0.1653)	***	-0.8832 (0.1652)	***	-0.8700 (0.1650)	***	0.4884 (0.2056)	**
d2001	0.1121 (0.0705)		-1.0108 (0.1790)	***	-1.0117 (0.1789)	***	-1.0002 (0.1787)	***	0.3980 (0.2205)	*
d2002	0.1582 (0.0738)	**	-1.0228 (0.1931)	***	-1.0248 (0.1929)	***	-1.0111 (0.1927)	***	0.2731 (0.2346)	
d2003	0.1010 (0.0762)		-1.2186 (0.2064)	***	-1.2199 (0.2063)	***	-1.2092 (0.2062)	***	-0.0125 (0.2468)	
d2004	0.1452 (0.0812)	*	-1.3768 (0.2221)	***	-1.3790 (0.2220)	***	-1.3679 (0.2220)	***	-0.2018 (0.2578)	
d2005	0.1589 (0.0879)	*	-1.2717 (0.2356)	***	-1.2743 (0.2353)	***	-1.2607 (0.2353)	***	-0.4193 (0.2728)	
d01Laval	-0.0577 (0.0865)		-0.1329 (0.1933)		-0.1344 (0.1932)		-0.1394 (0.1932)		1.6926 (0.2827)	***
d04MtIG	-0.0718 (0.0830)		-0.0141 (0.1820)		-0.0153 (0.1819)		-0.0208 (0.1819)		1.4065 (0.2634)	***
d07Concordia	0.4256 (0.1554)	***	0.8527 (0.3575)	**	0.8518 (0.3574)	**	0.8525 (0.3575)	**	-0.4596 (0.4994)	
d09UQAMG	-0.0827 (0.1015)		-0.3109 (0.2141)		-0.3137 (0.2138)		-0.3193 (0.2140)		1.7322 (0.3087)	***
d08SherbrookeG	-0.0890 (0.1174)		0.1583 (0.2581)		0.1566 (0.2580)		0.1502 (0.2580)		0.8857 (0.3740)	**
L2.nbarticlemoyan3			0.0302 (0.0292)		0.0352 (0.0226)					
L1.nbarticlemoyan3									0.0720 (0.0354)	**
L1.ltotsubvmoyan3wu			1.6286 (0.0943)	***	1.6284 (0.0943)	***	1.6292 (0.0943)	***		
L1.ltotcontmoyan3wu									0.1058 (0.0298)	***
L1.nb_loop_past									0.1953	**

Variable	FS_xtreg_art	FS_xtreg_subv	FS_xtreg_subv1	FS_xtreg_subv2	FS_xtreg_cont
					(0.0878)
L1.totcont_past					0.0000 ***
					(0.0000)
_cons	-0.2074	-12.5268 ***	-12.5368 ***	-12.5240 ***	-1.9700 ***
	(0.1715)	(1.0770)	(1.0763)	(1.0770)	(0.4832)
N	8772	8769	8769	8769	8771
N_g	1460	1460	1460	1460	1460
df_m	27	26	25	25	29
rank	28	27	26	26	30
chi2	5797.168	3257.078	3257.278	3256.009	892.7814
g_min	1	1	1	1	1
g_avg	6.008219	6.006164	6.006164	6.006164	6.007534
g_max	10	10	10	10	10
sigma	1.230234	3.106943	3.106536	3.106834	4.261584
sigma_u	0.852792	2.077302	2.076838	2.077305	3.066504
sigma_e	0.886691	2.310392	2.310262	2.310243	2.959333
rho	0.480519	0.447026	0.446943	0.447059	0.51778
r2_w	0.281491	0.299513	0.299529	0.299595	0.063057
r2_o	0.597065	0.247479	0.247404	0.246702	0.164723
r2_b	0.694161	0.111926	0.111883	0.111511	0.220868
p	0	0	0	0	3.6E-169

Tableau T-3: Régressions de première étape en Moindre Carrés Ordinaires pour les blocs B et C – breveter avec revendications et citations

Variable	reg_art		reg_subv		reg_subv1		reg_subv2		reg_cont	
L2.ltotsubvmoyan3	-0.1503 (0.0325)	***								
L1.ltotsubvmoyan3									-0.5472 (0.0781)	***
L2.sq_ltotsubvmoyan3	0.0134 (0.0029)	***								
L1.sq_ltotsubvmoyan3									0.0682 (0.0068)	***
L2.lmbetcent3	1.5541 (0.0704)	***	0.0622 (0.0650)				0.2062 (0.0464)	***	0.1804 (0.1104)	
codechair33	0.0053 (0.0339)		0.4364 (0.0459)	***	0.4387 (0.0454)	***	0.4393 (0.0462)	***	0.2285 (0.1385)	*
L1.age	-0.0451 (0.0233)	*	0.3275 (0.0471)	***	0.3292 (0.0435)	***	0.3238 (0.0471)	***	0.1736 (0.0768)	**
L1.sq_age	0.0022 (0.0011)	*	-0.0091 (0.0019)	***	-0.0092 (0.0018)	***	-0.0088 (0.0019)	***	-0.0069 (0.0035)	**
L1.nbarticlemoyan3wu	0.2101 (0.0788)	***								
L1.ltotcontmoyan3	0.0056 (0.0057)		0.1243 (0.0071)	***	0.1242 (0.0071)	***	0.1248 (0.0071)	***		
L1.lmcliquind23	0.1461 (0.0529)	***	-0.3938 (0.0759)	***	-0.3651 (0.0725)	***	-0.3998 (0.0756)	***	0.0824 (0.1567)	
L1.sq_lmcliquind23	-0.0071 (0.0021)	***	0.0151 (0.0030)	***	0.0139 (0.0029)	***	0.0153 (0.0030)	***	-0.0032 (0.0062)	
deg_app0	0.0953 (0.0094)	***	0.0452 (0.0210)	**	0.0456 (0.0208)	**	0.0510 (0.0209)	**	-0.1306 (0.0377)	***
L1.avg_pat_past	0.3958 (0.2155)	*	0.1483 (0.1066)		0.1228 (0.1093)		0.1921 (0.1269)		-1.1109 (0.4526)	**
nano_excl	0.0936 (0.1021)		0.5147 (0.2066)	**	0.4363 (0.2064)	**	0.5266 (0.2058)	**	-0.4424 (0.5775)	
bio_excl	-0.1840 (0.0427)	***	-0.0549 (0.0856)		-0.0636 (0.0848)		-0.0642 (0.0855)		0.3045 (0.2255)	
d1997	0.3241 (0.0674)	***	-3.0472 (0.1921)	***	-3.0382 (0.1875)	***	-3.0413 (0.1925)	***	0.2595 (0.1612)	
d1998	0.1900 (0.0743)	**	-0.6181 (0.1419)	***	-0.6493 (0.1412)	***	-0.6021 (0.1424)	***	0.3271 (0.2105)	
d1999	0.3529 (0.0919)	***	-0.6330 (0.1486)	***	-0.6695 (0.1474)	***	-0.6091 (0.1486)	***	0.2973 (0.2519)	
d2000	0.2769 (0.0959)	***	-1.0849 (0.1731)	***	-1.1592 (0.1726)	***	-1.0553 (0.1729)	***	0.4852 (0.2874)	*
d2001	0.3052 (0.0981)	***	-1.2892 (0.1868)	***	-1.3125 (0.1852)	***	-1.2605 (0.1864)	***	0.4045 (0.3024)	
d2002	0.4544 (0.1028)	***	-1.2729 (0.1960)	***	-1.3367 (0.1938)	***	-1.2353 (0.1953)	***	0.1535 (0.3085)	
d2003	0.3214 (0.0995)	***	-1.5008 (0.2122)	***	-1.5527 (0.2096)	***	-1.4709 (0.2119)	***	-0.0993 (0.3168)	
d2004	0.4142 (0.1060)	***	-1.7382 (0.2356)	***	-1.7969 (0.2324)	***	-1.7099 (0.2355)	***	-0.2975 (0.3293)	
d2005	0.4471 (0.1118)	***	-1.7771 (0.2492)	***	-1.8436 (0.2461)	***	-1.7428 (0.2491)	***	-0.4710 (0.3427)	
d01Laval	-0.0681 (0.1099)		0.0221 (0.1225)		0.0313 (0.1212)		0.0050 (0.1224)		1.9389 (0.3424)	***
d04MtIG	-0.1092 (0.1074)		0.1013 (0.1163)		0.1067 (0.1148)		0.0810 (0.1163)		1.5474 (0.3041)	***
d07Concordia	0.2699 (0.2168)		0.7614 (0.3093)	**	0.7911 (0.2923)	***	0.7570 (0.3101)	**	-0.0635 (0.3198)	
d09UQAMG	0.0277 (0.1395)		-0.1997 (0.1366)		-0.1768 (0.1347)		-0.2146 (0.1368)		2.4038 (0.3426)	***
d08SherbrookeG	-0.1110 (0.1195)		0.2050 (0.1501)		0.2406 (0.1501)		0.1816 (0.1499)		1.0860 (0.4052)	***
L2.nbarticlemoyan3	0.0823 (0.0261)	***	0.0985 (0.0186)	***						
L1.ltotsubvmoyan3wu	1.4189 (0.1422)	***	1.4534 (0.1405)	***	1.4220 (0.1423)	***				
L1.ltotcontmoyan3wu	0.0935	***								

Variable	reg_art	reg_subv	reg_subv1	reg_subv2	reg_cont
	(0.0357)				
L1.nb_loop_past	0.6847 ***				
	(0.1413)				
L1.totcont_past	0.0000 *				
	(0.0000)				
_cons	-0.1888	-7.7424 ***	-8.1453 ***	-7.7458 ***	-2.4424 ***
	(0.2373)	(1.5895)	(1.5591)	(1.5909)	(0.6153)
Statistics					
N	8771	8769	8929	8769	8769
df_m	28	27	26	26	29
rank	29	28	27	27	30
ll	-15153	-21338.3	-21780.3	-21344.1	-25573
ll_0	-19082.5	-22652	-23139.6	-22652	-26500.9
aic	30363.97	42732.55	43614.63	42742.27	51205.95
bic	30569.26	42930.76	43806.25	42933.4	51418.32
r2	0.591815	0.258906	0.26248	0.257916	0.190742





Variable	FS_xtreg_art		FS_xtreg_subv		FS_xtreg_subv1		FS_xtreg_subv2		FS_xtreg_cont	
									(0.0311)	
L1.nb_loop_past									0.2179	**
									(0.0856)	
L1.totcont_past									0.0000	***
									(0.0000)	
_cons	-0.6246	***	-12.2805	***	-12.6771	***	-12.2806	***	-2.6286	***
	(0.1789)		(1.0757)		(1.0601)		(1.0757)		(0.4824)	
N	8771		8769		8929		8769		8769	
N_g	1460		1460		1469		1460		1460	
n_clust										
df_m	28		27		26		26		29	
rank	29		28		27		27		30	
chi2	5014.602		3261.467		3378.607		3260.78		933.6449	
g_min	1		1		1		1		1	
g_avg	6.007534		6.006164		6.078285		6.006164		6.006164	
g_max	10		10		10		10		10	
sigma	1.251955		3.091167		3.072871		3.091115		4.265585	
sigma_u	0.858731		2.053439		2.001391		2.053504		3.07869	
sigma_e	0.911028		2.310564		2.331732		2.310435		2.952437	
rho	0.470475		0.441284		0.424204		0.441327		0.520924	
r2_w	0.240258		0.299407		0.303705		0.299502		0.067117	
r2_o	0.574095		0.249151		0.253084		0.248445		0.16885	
r2_b	0.672746		0.115829		0.11701		0.115502		0.222338	
p	0		0		0		0		8.80E-178	

## ANNEXE U – Régressions de la première étape pour la phase 2 (blocs D, E et F)

Tableau U-1: Régressions de première étape en Moindres Carrés Ordinaires pour le bloc D – nombre de brevets

Variable	FS_article		FS_subv		FS_subv1		FS_subv2		FS_cont	
L2.ltotsubvmoyan3	-0.5411 (0.1282)	***							0.4028 (0.0589)	***
L2.sq_ltotsubvmoyan3	0.0478 (0.0122)	***								
L2.ldmbetcent3	1.8558 (0.2269)	***	-0.1554 (0.1638)				-0.1554 (0.1638)			
L1.nbarticlemoyan3wu	-0.6945 (0.4894)									
L1.ltotcontmoyan3	-0.0351 (0.0232)		0.1282 (0.0230)	***	0.1269 (0.0233)	***	0.1282 (0.0230)	***		
L2.lmcliqind23	0.0689 (0.2383)		-0.4156 (0.3073)		-0.4888 (0.2947)	*	-0.4156 (0.3073)		-0.2620 (0.6044)	
L2.sq_lmcliqind23	-0.0025 (0.0096)		0.0157 (0.0120)		0.0188 (0.0115)		0.0157 (0.0120)		0.0073 (0.0238)	
L1.deg_app0	0.2168 (0.0495)	***	0.0174 (0.0849)		0.0076 (0.0825)		0.0174 (0.0849)		0.0275 (0.1803)	
L1.avg_pat_past	-0.0300 (0.1096)		0.1517 (0.0872)	*	0.1464 (0.0893)		0.1517 (0.0872)	*	-0.2400 (0.4397)	
L1.o_patpast	0.4238 (0.1335)	***	-0.2147 (0.1563)		-0.2174 (0.1576)		-0.2147 (0.1563)		-0.4167 (0.3649)	
L1.nbloop0	-0.1117 (0.0478)	**	0.0972 (0.0526)	*	0.0965 (0.0539)	*	0.0972 (0.0526)	*	-0.1758 (0.2889)	
codechair33	0.0736 (0.1446)		0.2883 (0.1539)	*	0.2904 (0.1541)	*	0.2883 (0.1539)	*	0.2772 (0.3504)	
age	-0.0957 (0.1414)		0.0057 (0.2057)		-0.0049 (0.2074)		0.0057 (0.2057)		-0.2170 (0.4123)	
sq_age	0.0024 (0.0063)		0.0045 (0.0082)		0.0048 (0.0082)		0.0045 (0.0082)		0.0072 (0.0164)	
d1997	0.7426 (0.3850)	*	-3.2645 (0.6837)	***	-3.2296 (0.6878)	***	-3.2645 (0.6837)	***	-1.2402 (1.1147)	
d1998	0.1963 (0.4332)		-0.7724 (0.5049)		-0.7660 (0.5112)		-0.7724 (0.5049)		0.9907 (1.0700)	
d1999	0.5878 (0.5737)		-1.7415 (0.4950)	***	-1.7187 (0.5044)	***	-1.7415 (0.4950)	***	0.3502 (0.9578)	
d2000	0.4294 (0.5470)		-1.3597 (0.4754)	***	-1.3467 (0.4815)	***	-1.3597 (0.4754)	***	0.0402 (1.0168)	
d2001	0.6103 (0.5406)		-2.8325 (0.6338)	***	-2.8287 (0.6362)	***	-2.8325 (0.6338)	***	0.5156 (1.1791)	
d2002	0.8676 (0.5714)		-2.6361 (0.6689)	***	-2.5938 (0.6728)	***	-2.6361 (0.6689)	***	-0.2276 (1.2011)	
d2003	0.5268 (0.5431)		-2.9754 (0.7255)	***	-2.9539 (0.7279)	***	-2.9754 (0.7255)	***	-0.7436 (1.3301)	
d2004	0.4282 (0.4916)		-3.3912 (0.8304)	***	-3.3551 (0.8280)	***	-3.3912 (0.8304)	***	-1.3469 (1.3962)	
d2005	0.5608 (0.5586)		-3.4146 (0.8177)	***	-3.3989 (0.8162)	***	-3.4146 (0.8177)	***	-0.3885 (1.5096)	
d01Laval	-0.2472 (0.4497)		0.4616 (0.3527)		0.5175 (0.3441)		0.4616 (0.3527)		-0.4350 (1.1213)	
d04MtlG	-0.7717 (0.4632)	*	0.3286 (0.3550)		0.3489 (0.3513)		0.3286 (0.3550)		0.9497 (1.0180)	
d07Concordia	-1.5331	***	0.8879		0.8780		0.8879		-3.2503	***

Variable	FS_article	FS_subv	FS_subv1	FS_subv2	FS_cont	
	(0.5733)	(0.6384)	(0.6338)	(0.6384)	(0.8155)	
d09UQAMG	-0.2243	0.4906	0.5389	0.4906	2.9659	***
	(0.5131)	(0.3800)	(0.3809)	(0.3800)	(1.0844)	
d08SherbrookeG	-1.1754 **	1.2114 ***	1.2250 ***	1.2114 ***	-0.2465	
	(0.5457)	(0.4162)	(0.4185)	(0.4162)	(1.4323)	
L2.nbarticlemoyan3		0.1671 ***	0.1336 ***	0.1671 ***		
		(0.0478)	(0.0346)	(0.0478)		
L1.nbarticlemoyan3					0.1604	
					(0.2209)	
L1.ltotsubvmoyan3wuu		2.2678 ***	2.2736 ***	2.2678 ***		
		(0.4762)	(0.4768)	(0.4762)		
L1.ltotcontmoyan3wuu					0.0447	
					(0.1397)	
L1.nb_loop_past					0.5615	***
					(0.1824)	
L1.totcont_past					0.0000	***
					(0.0000)	
L1.sq_nbartic~3					-0.0032	
					(0.0092)	
_cons	1.6442	-15.2962 ***	-15.2110 ***	-15.2962 ***	2.2016	
	(1.2910)	(5.6293)	(5.6190)	(5.6293)	(3.1611)	
N	526	526	526	526	526	
df_m	28	27	26	27	29	
rank	29	28	27	28	30	
ll	-1090.32	-1211.98	-1212.36	-1211.98	-1559.94	
ll_0	-1353.5	-1321.16	-1321.16	-1321.16	-1637.89	
aic	2238.639	2479.964	2478.712	2479.964	3179.889	
bic	2362.333	2599.393	2593.875	2599.393	3307.848	
r2	0.632379	0.339752	0.338813	0.339752	0.256494	

Tableau U-2: Régressions de première étape en Moindres Carrés Ordinaires pour les blocs E et F – nombre de revendications et citations

Variable	FS_article		FS_subv		FS_subv1		FS_subv2		FS_cont	
L2.ltotsubvmoyan3	-0.6790 (0.1737)	***								
L1.ltotsubvmoyan3									0.5812 (0.1111)	***
L2.sq_ltotsubvmoyan3	0.0582 (0.0162)	***								
L2.ldmbetcent3	1.9032 (0.2454)	***	0.0494 (0.1718)				0.3525 (0.1405)	**		
L1.nbarticlemoyan3wu	-0.8641 (0.6738)									
L1.ltotcontmoyan3	0.0259 (0.1570)		0.1470 (0.0901)		0.1455 (0.0898)		0.1170 (0.0917)			
L1.sq_ltotcontmoyan3	-0.0048 (0.0146)		-0.0004 (0.0068)		-0.0004 (0.0068)		0.0024 (0.0071)			
L1.lmcliquind23	0.3947 (0.2486)		-0.4025 (0.2891)		-0.4003 (0.2803)		-0.4081 (0.2853)		-0.2201 (0.6728)	
L1.sq_lmcliquind23	-0.0168 (0.0099)	*	0.0188 (0.0115)		0.0186 (0.0111)	*	0.0192 (0.0113)	*	0.0033 (0.0263)	
deg_app	-0.2458 (0.1719)		0.2791 (0.2489)		0.3185 (0.2483)		0.2504 (0.2454)		-0.2300 (0.4548)	
nbloop	0.0026 (0.0536)		0.2345 (0.0585)	***	0.2242 (0.0568)	***	0.2355 (0.0591)	***	-0.1942 (0.2187)	
propuninv	-0.0678 (0.4840)		1.0595 (0.5515)	*	1.0692 (0.5392)	**	1.0559 (0.5638)	*	-0.0999 (1.0864)	
excl_nano_pat	0.3167 (0.3260)		-0.2778 (0.3432)		-0.2811 (0.3387)		-0.2968 (0.3449)		-0.2817 (0.8907)	
excl_bio_pat	-0.6427 (0.3788)	*	-0.5638 (0.2864)	*	-0.5638 (0.2848)	**	-0.6282 (0.2854)	**	0.8167 (0.6506)	
codechair33	0.1363 (0.1655)		0.1828 (0.1613)		0.1718 (0.1634)		0.2141 (0.1600)		0.1133 (0.3809)	
age	0.0710 (0.2005)		0.1419 (0.2642)		-0.0032 (0.2414)		0.1554 (0.2656)		-0.5535 (0.4687)	
sq_age	-0.0020 (0.0085)		-0.0015 (0.0104)		0.0037 (0.0095)		-0.0019 (0.0104)		0.0221 (0.0184)	
d1997	0.9962 (0.4526)	**	-2.6548 (0.7384)	***	-2.6420 (0.7363)	***	-2.6153 (0.7519)	***	0.2546 (1.2547)	
d1998	0.0996 (0.5176)		-0.9615 (0.5620)	*	-0.8738 (0.5392)		-0.9208 (0.5820)		1.4345 (1.1852)	
d1999	0.8723 (0.7165)		-2.1028 (0.6709)	***	-1.9328 (0.6506)	***	-2.0360 (0.6928)	***	2.2412 (1.2589)	*
d2000	0.7254 (0.6663)		-1.5163 (0.5884)	**	-1.3878 (0.5756)	**	-1.4803 (0.6120)	**	1.4337 (1.3057)	
d2001	0.7531 (0.6320)		-3.0602 (0.7564)	***	-2.9124 (0.7489)	***	-3.0483 (0.7759)	***	2.1503 (1.4769)	
d2002	0.8929 (0.7112)		-2.4342 (0.7237)	***	-2.3059 (0.7210)	***	-2.3333 (0.7447)	***	1.9534 (1.4952)	
d2003	0.5075 (0.6471)		-3.1640 (0.8846)	***	-3.0423 (0.8831)	***	-3.1140 (0.9006)	***	1.0782 (1.4793)	
d2004	0.4773 (0.6445)		-3.7024 (0.9889)	***	-3.5994 (0.9864)	***	-3.6570 (0.9925)	***	0.5130 (1.6473)	
d2005	1.1278 (0.7010)		-3.1725 (1.0142)	***	-3.0426 (1.0074)	***	-3.1361 (1.0251)	***	0.9809 (1.7985)	
d01Laval	-0.6448 (0.5913)		0.3767 (0.4084)		0.3653 (0.4024)		0.4190 (0.4151)		0.5037 (1.1996)	
d04MtIG	-1.1159 (0.6086)	*	0.4918 (0.4030)		0.5620 (0.4011)		0.4327 (0.4024)		0.7538 (1.0857)	
d07Concordia	-1.9027 (0.7694)	**	1.4131 (0.7661)	*	1.3559 (0.7611)	*	1.3185 (0.7566)	*	-3.0852 (0.8526)	***
d09UQAMG	-0.5882 (0.6261)		0.6882 (0.4662)		0.7077 (0.4666)		0.6940 (0.4779)		2.5343 (1.2495)	**
d08SherbrookeG	-1.6530 (0.6928)	**	1.1034 (0.4831)	**	1.0703 (0.4864)	**	1.0309 (0.4958)	**	-0.3292 (1.6143)	
L2.nbarticlemoyan3			0.1272 (0.0466)	***	0.1410 (0.0374)	***				
L1.ltotsubvmoyan3wu			2.7019	***	2.6205	***	2.7059	***		

Variable	FS_article	FS_subv	FS_subv1	FS_subv2	FS_cont
		(0.5613)	(0.5542)	(0.5690)	
L1.totcontmoyan3wu					-0.1365 (0.1954)
L1.nb_loop_past					0.3840 ** (0.1627)
L1.totcont_past					0.0000 *** (0.0000)
_cons	2.3114 (1.7858)	-22.6497 *** (6.5086)	-20.9452 *** (6.2467)	-22.7057 *** (6.5987)	2.4188 (3.8787)
N	419	419	422	419	422
df_m	30	29	28	28	28
rank	31	30	29	29	29
ll	-902.62618	-965.65184	-972.85551	-968.07718	-1257.9707
ll_0	-1102.3488	-1062.3642	-1068.6272	-1062.3642	-1317.254
aic	1867.2524	1991.3037	2003.711	1994.1544	2573.9415
bic	1992.4264	2112.4398	2121.0162	2111.2526	2691.2466
r2	0.61454464	0.36974723	0.36485013	0.36240853	0.2449441

## ANNEXE V – Extraits procéduraux de forage de texte et de méthodes économétriques

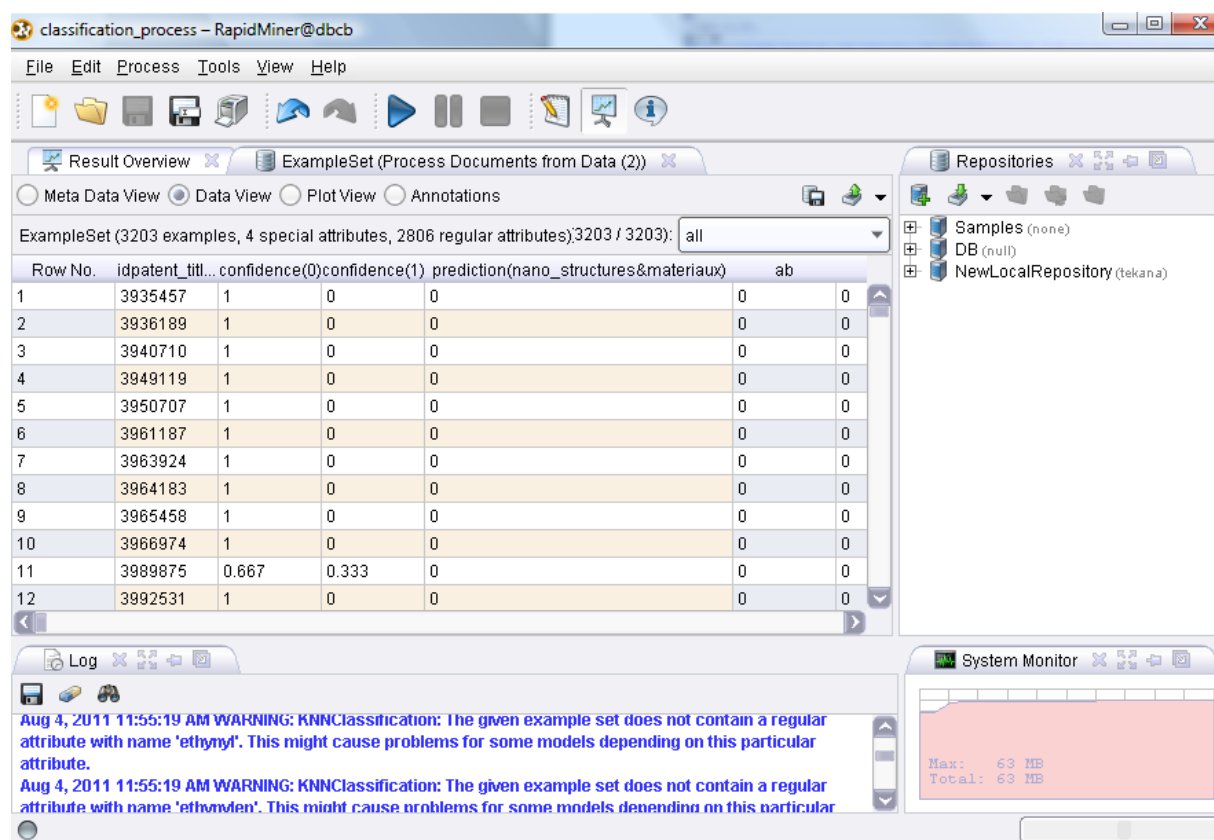


Figure V-1: Extrait des résultats de la classification dans la catégorie de nanostructures et nanomatériaux

```

65
66 * FIXED EFFECTS
67
68 xtreg lsum_nb_claims ///
69     ltotsubvmoyan3 ltotcontmoyan3 ldmбетcentind3 lmcliquind23 ///c
70     d01Laval d02McGill d04Mt1G d07Concordia d09UQAM ///
71     d1996 d1997 d1998 d1999 d2000 d2001 d2002 d2003 d2004 d2005 ///
72     codechair33 age deg_app loop_pat_app if chercheur_inventeur==1 & year>=1996, fe
73 estimates store fixed_nb_claims_xtreg
74
75 * RANDOM EFFECTS
76
77 xtreg lsum_nb_claims ///
78     ltotsubvmoyan3 ltotcontmoyan3 ldmбетcentind3 lmcliquind23 ///
79     d01Laval d02McGill d04Mt1G d07Concordia d09UQAM ///
80     d1996 d1997 d1998 d1999 d2000 d2001 d2002 d2003 d2004 d2005 ///
81     codechair33 age deg_app loop_pat_app if chercheur_inventeur==1 & year>=1996, re
82 estimates store random_nb_claims_xtreg
83
84 * HAUSMAN TEST: To determine which is more appropriate, fixed or random effects model.
85 *****HO: random effects estimation is equivalent to fixed effects estimation
86 *****H1: random effects estimation is not acceptable and fixed effects must be taken.
87
88 hausman fixed_nb_claims_xtreg random_nb_claims_xtreg
89 estimates store H1
90
91 *BREUSCH-PAGAN TEST:
92 *****HO: no specific random effects
93 *****H1: specific random effects present and
94 xtreg lsum_nb_claims ///
95     ltotsubvmoyan3 ltotcontmoyan3 ldmбетcentind3 lmcliquind23 ///
96     d01Laval d02McGill d04Mt1G d07Concordia d09UQAM ///
97     d1996 d1997 d1998 d1999 d2000 d2001 d2002 d2003 d2004 d2005 ///
98     codechair33 age deg_app loop_pat_app if chercheur_inventeur==1 & year>=1996, re
99 xttest0
100 estimates store BP1
101
102
103
104

```

Figure V-2: Tests d'Hausman et de Breusch-Pagan dans Stata



```

296
297 ***First stage: REG on contracts***
298
299 reg L.ltotcontmoyan3
300     L2.lmbetcentind3 codechair33 L.age L.sq_age L.ltotcontmoyan3wuu L.nb_loop_past L.totcont_past    /// var inst
301     L.nbarticlemoyan3 L2.ltotsubvmoyan3    ///
302     L2.lmcliquind23 L2.sq_lmcliquind23    /// exogenes
303     L.deg_app0 L.avg_pat_past L.o_patpast L.nbloop0    /// autres e
304     `years'    ///
305     `clusters'    ///
306     `conditions'    ///
307     , vce(cluster id)
308
309 predict L_cont_res, res
310
311 ***Test for effect of res***
312
313
314 tobit lnbpast    ///
315     L_nbarticlemoyan3 L2_ltotsubvmoyan3    ///
316     L2_lmcliquind23 L2_sq_lmcliquind23    /// exogenes
317     L_deg_app0 L_avg_pat_past L_o_patpast L_nbloop0    /// autres e
318     L_ltotcontmoyan3 L_cont_res    ///
319     `years'    ///
320     `clusters'    ///
321     `conditions'    ///
322     , vce(cluster id) ll
323 estat ic
324 estimates store HSM_tobit_cont
325 test L_cont_res
326

```

Figure V-3: Test de Durbin-Wu-Hausman dans Stata

## ANNEXE W – Variables employées dans les régressions

Tableau W-1: Liste des variables employées

catégorie	Variable	Type	Description
<b>financement</b>	ltotsubvmoyan3	moyenne	moyenne du total des subventions reçues par le chercheur sur 3 ans
	sq_ltotsubvmoyan3	moyenne	carré de la moyenne du total des subventions reçues par le chercheur sur 3 ans
	totsubvmoyan3wuu	moyenne	moyenne du total des subventions reçues par l'ensemble de l'université d'affiliation du chercheur sur 3 ans
	totsubv_past	total	total des subventions dans le passé
	ltotcontmoyan3	moyenne	moyenne du total des contrats reçus par chercheur le sur 3 ans
	sq_ltotcontmoyan3	moyenne	carré moyenne du total des contrats reçus par chercheur le sur 3 ans
	totcontmoyan3wuu	moyenne	moyenne du total des contrats reçues par l'ensemble de l'université d'affiliation du chercheur sur 3 ans
	totcont_past	total	total des contrats reçus dans le passé
<b>articles</b>	nbarticlemoyan3	moyenne	moyenne du nombre d'articles publiés par le chercheur sur 3 ans
	sq_nbarticlemoyan3	moyenne	carré de la moyenne du nombre d'articles publiés par le chercheur sur 3 ans
	nbarticlemoyan3wuu	moyenne	moyenne du nombre d'articles publiés de l'ensemble de l'université d'affiliation du chercheur sur 3 ans
	deg_app	moyenne	degré d'application des connaissances associées à la recherche effectuée par le chercheur
	deg_app0	moyenne	degré d'application des connaissances associées à la recherche effectuée par le chercheur avec 0 par défaut
	deg_appmoyan3	moyenne	moyenne du degré d'application sur une période de 3 ans avec la valeur par défaut 0
	deg_appmoyan4	moyenne	moyenne du degré d'application sur une période de 4 ans avec la valeur par défaut 1
	nbart_cumul	nombre	nombre cumulatif d'articles pour le chercheur à l'année
	nbart_past	nombre	nombre d'articles passés pour le chercheur à l'année
<b>chercheur</b>	age	nombre	temps depuis la première publication du chercheur
	sq_age	nombre	carré du temps depuis la première publication du chercheur
	codechair33	catégorie (1,2,3)	Le type de chaire que possède le chercheur
	chercheur_inventeur	dummy	si le chercheur a au moins un brevet
	chercheur_post_invention	dummy	moment après que le chercheur ait eu son premier brevet

catégorie	Variable	Type	Description
<b>réseaux</b>	lmcliqind23	nombre	densité égocentrique du chercheur
	sq_lmcliqind23	nombre	carré de la densité égocentrique du chercheur
	ldmbetcent3	nombre	degré d'intermédiarité du chercheur
<b>variables interactives</b>	deg_app_nbartcile	interactive	variable interactive de deg_app0*nbarticle0
	ltot_subv_cont_moyan3	interactive	variable interactive de ltotsubvmoyan3*L.totcontmoyan3
	ltotcontmoyan3_age	interactive	variable interactive de ltotcontmoyan3*age
	ldm_betcent_cliqind23	interactive	variable interactive de L3.ldmbetcentind3*lmcliqind23
	ldm_betcent_nbarticle	interactive	variable interactive de L3.ldmbetcentind3*nbarticle0
<b>varia</b>	years (d1996-d2005)	dummies	à l'année t ou non
	clusters (dConcordia-dUQAM)	dummies	université d'appartenance du chercheur avec McGill comme université de référence
	nano_excl	dummy	si le chercheur fait de la recherche en nanotechnologie
	bio_excl	dummy	si le chercheur fait de la recherche en biotechnologie
<b>préfixes</b>	L. ou L_*	-	Lag d'une année sur la variable spécifiée
	L2. ou L2_	-	Lag de deux années sur la variable spécifiée
	L3. ou L3_	-	Lag de trois années sur la variable spécifiée

Tableau W-2: Liste des variables employées (suite)

catégorie	Variable	Type	Description
brevets	arith_avg_pat_10yrs	moyenne	moyenne arithmétique du nombre de brevets d'un chercheur sur les 10 dernières années
	aucun_brevet_en_boucle	dummy	pour cette année-chercheur, il n'y a pas de brevets en boucle
	avg_age_boucle	moyenne	moyenne d'age de la boucle du financement au brevetage
	avg_age_boucle30	moyenne	moyenne d'age de la boucle du financement au brevetage avec les brevets hors boucle ayant un age par défaut de 30
	avg_age_boucle40	moyenne	moyenne d'age de la boucle du financement au brevetage avec les brevets hors boucle ayant un age par défaut de 40
	avg_pat_past	moyenne	moyenne du nombre de brevets don't le chercheur est inventeur dans le passé (10 ans ou age)
	bioinformatics_pat	dummy	si le brevet est dans ce sous-domaine
	c_arith_avg_pat_10yrs	moyenne	moyenne arithmétique corrigée du nombre de brevets d'un chercheur sur les 10 dernières années
	cell_tissue_engineering_pat	dummy	si le brevet est dans ce sous-domaine
	devices_electronics_pat	dummy	si le brevet est dans ce sous-domaine
	dna_rna_pat	dummy	si le brevet est dans ce sous-domaine
	en_et_hors_boucle	dummy	pour cette année-chercheur, il y a des brevets hors-boucle et dans-boucle
	excl_bio_pat	dummy	brevets exclusivement biotechnologiques
	excl_nano_pat	dummy	brevets exclusivement nanotechnologiques
	gene_rna_vectors_pat	dummy	si le brevet est dans ce sous-domaine
	geom_avg_pat_10yrs	moyenne	moyenne arithmétique du nombre de brevets d'un chercheur sur les 10 dernières années
	geom_avg_pat_10yrs0	moyenne	moyenne arithmétique du nombre de brevets d'un chercheur sur les 10 dernières années avec la valeur par défaut 0
	ln_nb_patents	nombre	logarithme naturel du nombre de brevets
	medicine_biotech_pat	dummy	si le brevet est dans ce sous-domaine
	metrology_processes_pat	dummy	si le brevet est dans ce sous-domaine
	mov_avg_pat_3yrs	moyenne	moyenne mobile du nombre de brevets sur 3 ans (Lag +year+ forward)
	mov_avg_pat_5yrs	moyenne	moyenne mobile du nombre de brevets sur 5 ans(Lag2+Lag +year+ forward+ forward2)
	mult_pat_0_1	dummy	l'année-chercheur a 1 brevet ou plus
	mult_pat_0_1_2	ordinal	l'année chercheur a 0 brevet(0), 1 brevet(1) ou 2 brevets ou plus(2)
	o_pat_past	ordinal	le chercheur a eu 0(0), 1(1) ou 2 brevets ou plus (2) dans le passé
	mult_pat_per_year_past	dummy	le chercheur a déjà eu 2 brevets ou plus à une même année dans le passé
	nano_bio_pat	dummy	l'année-chercheur a autant des brevets bio que nano
	nbloop0	nombre	nombre de brevets en boucle avec la valeur par défaut 0
	nb_loop_past	total	nombre de brevets issus de boucles inventés par le chercheur dans le passé
	nb_bio_exclusif	nombre	nombre de brevets exclusivement bio à l'année
	nb_nano_bio	nombre	nombre de brevets en nanobiotech à l'année
	nb_nano_exclusif	nombre	nombre de brevets exclusivement nano à l'année
	nb_patents0	nombre	nombre de brevets avec la valeur par défaut 0

catégorie	Variable	Type	Description
	pat_past_0_1	dummy	has si le checheur a déjà eu au moins un brevet dans le passé
	process_bio_techniques_pat	dummy	si le brevet est dans ce sous-domaine
	propuninv	moyenne	proportion d'universitaires dans l'équipe de brevetage
	proteins_molecules_pat	dummy	si le brevet est dans ce sous-domaine
	structures_materials_pat	dummy	si le brevet est dans ce sous-domaine
	sum_bioinformatics	nombre	somme de brevets dans ce sous-domaine
	sum_cell_tissue_engineering	nombre	somme de brevets dans ce sous-domaine
	sum_devices_electronics	nombre	somme de brevets dans ce sous-domaine
	sum_dna_rna	nombre	somme de brevets dans ce sous-domaine
	sum_gene_rna_vectors	nombre	somme de brevets dans ce sous-domaine
	sum_medicine_biotech	nombre	somme de brevets dans ce sous-domaine
	sum_metrology_processes	nombre	somme de brevets dans ce sous-domaine
	sum_process_bio_techniques	nombre	somme de brevets dans ce sous-domaine
	sum_proteins_molecules	nombre	somme de brevets dans ce sous-domaine
	sum_structures_materials	nombre	somme de brevets dans ce sous-domaine
	tous_brevets_en_boucle	dummy	pour cette année-chercheur, tous les brevets sont en boucle